

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 4, 2003, 09:32:34 ; Search time 2444 Seconds
(without alignments)
4366.950 Million cell updates/sec

Title: US-09-802-285A-2
Perfect score: 3494
Sequence: 1 MTTFKRIIFAVIALSSG.....KGKLNLTITNGKQQLVLP 659

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09802285/runat_28072003_190511_9730/app_query.fasta_1.839
-DB=EST -QPWT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=DCT -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09802285@cgn_1_1906 @runat_28072003_190511_9730 -NCFU=6 -ICFU=3
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:.*
1: em_estba:.*
2: em_esthum:.*
3: em_estin:.*
4: em_estmu:.*
5: em_estov:.*
6: em_estpl:.*
7: em_estro:.*
8: em_htc:.*
9: gb_est1:.*
10: gb_est2:.*
11: gb_hic:.*
12: gb_est3:.*
13: gb_est4:.*
14: gb_est5:.*
15: em_estfun:.*
16: em_estom:.*
17: gb_gss:.*
18: em_gss_hum:.*
19: em_gss_inv:.*
20: em_gss_pln:.*
21: em_gss_vrt:.*
22: em_gss_fun:.*
23: em_gss_nam:.*
24: em_gss_mus:.*
25: em_gss_other:.*
26: em_gss_pro:.*
27: em_gss_rod:.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	
1	109.5	3.1	2547	11	AK010256	AK010256 Mus muscu	
2	107.5	3.1	2637	11	AK011146	AK011146 Mus muscu	
3	99.5	2.8	957	12	BG566122	BG566122 602582746	
C	4	98.5	2.8	580	10	AW942324	LD14783.3
	5	97.5	2.8	641	13	BJ420555	BJ420555 BJ420555
6	97.5	2.8	655	13	BJ424970	BJ424970 BJ424970	
7	97.5	2.8	659	13	BJ421576	BJ421576 BJ421576	
8	97.5	2.8	663	13	BJ411853	BJ411853 BJ411853	
9	97.5	2.8	815	14	BQ752776	WHE4119 A	
10	95.5	2.7	1050	11	AK006271	Mus muscu	
11	95	2.7	671	14	BQ744898	BQ744898 946112B02	
12	94	2.7	1001	17	CNS06JUZ	T7 end of	
13	93	2.7	756	13	BM413117	BM413117 EST587444	
14	93	2.7	762	13	BM165623	BM165623 EST568146	
15	93	2.7	780	13	BF647560	BF647560 603278139	
16	93	2.7	953	12	BF686272	BF686272 602142733	
C	17	92.5	2.6	625	9	AL649582	AL649582
	18	92.5	2.6	716	13	BM161568	BM161568 EST564091
19	92.5	2.6	734	13	BI889777	BI889777 ZF637-2-0	
C	20	92.5	2.6	966	12	BG120749	BG120749 602348507
21	92.5	2.6	2587	11	AK010578	Mus muscu	
22	92	2.6	809	9	AL669439	AL669439	
23	92	2.6	2588	11	AY108732	Zea mays	
24	91.5	2.6	953	14	BQ891785	AGENCOURT	
C	25	91.5	2.6	966	12	BG114631	602315418
	26	91.5	2.6	1072	10	BE378651	601237043
C	27	91	329	17	AZ769390	IM0570H03	
28	91	2.6	611	17	BH670936	BOHTE55TR	
29	91	2.6	668	12	BF294532	006PB03	
30	91	2.6	883	14	BQ687587	AGENCOURT	
31	91	2.6	947	17	CNS015IM	Drosophil	
C	32	91	1028	17	CNS048YA	AGENCOURT	
33	90.5	2.6	568	17	BH371376	AG-ND-125	
34	90.5	2.6	584	14	BQ450998	PFEST0aB0	
35	90.5	2.6	743	14	BQ917010	QHB20A01	
C	36	90.5	2.6	1200	14	BQ890483	AGENCOURT
	37	90	2.6	621	13	BJ338430	BU338430
38	90	2.6	713	9	AU101720	AU101720	
39	90	2.6	875	17	BH155081	ENTQ23TF	
40	90	2.6	929	13	BI147055	602913334	
C	41	90	1036	17	CNS07B0G	T3 end of	
42	90	2.6	1896	11	AK003169	Mus muscu	
43	89.5	2.6	586	13	BM317522	LG1_332.F	
44	89.5	2.6	625	17	AQ034246	CpG0543A	
45	89.5	2.6	658	14	BQ519807	rd01g12.Y	

ALIGNMENTS

RESULT 1
AK010256
LOCUS
DEFINITION
Mus musculus ES cells cDNA, RIKEN full-length enriched library,
expressed, full insert sequence.
ACCESSION
AK010256
VERSION
AK010256.1 GI:12845561
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (strain:C57BL/6J) ES cells cDNA to mRNA,
clone:lib:RIKEN full-length enriched mouse cDNA library
clone:2400002G22.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.


```

Query Match: 3.08% Indels: 220
DB: 11 Gaps: 40
US-09-802-285A-2 (1-659) x AK0111146 (1-2637)

QY 18 SerSerGlyAsnIleLeuAla-GlnSerSerSerIleThrArgLysAspPheAspHisI1 37
DB 561 AGCGAGCGGTATCTGGCACTCCAGTCCGACATATGATGGCGAGGCCATCGAGAGAGT 620
QY 37 eAsnLeuGluTyrSerGlyLeuGluLysValAsnLysAlaAlaGlyAsnTyrAs 57
DB 621 C-----TATGCGGATGTCTATGTATCGGCCCAACCAATAGAAAACGGCTCTATTA 671
QY 57 pAspAlaAlaLysAlaLeuLeuAlaTyrTyrArgGluLysSerLysAlaArgGluProAs 77
DB 672 TGAT-----ATGTACTTGAAGAAAGGGCGGTGTCAGCAATGA 710
QY 77 pPheSerAsnAlaGlu-----LysProAlaAspIleArgGlnProIleAspLys 93
DB 711 CTCTCTCCCTGGAAACTTTGTGCAAGAAATCATTAAGAGAAACAACCTTTGGAAG 770
QY 93 sValThrArgGluMetAlaAspLysAlaLeuValHisGlnPheGlnProHisLysGlyTy 113
DB 771 ATTG-----GAAGTTAAAGAGCAACATTACTAGAAATGTTTAAG-----TA 812
QY 113 rClyTyrPheAsp-----TyrGlyLysAspIleAsn-----Tr 124
DB 813 CAACAGTTCAGTCCGGATATTGAATGAGAAAGTGAATATCTCCGACCACTGTCTTA 872
QY 124 pGlnMetTrpProValLysAsp-----AsnGluValArg----- 135
DB 873 CAGATGTGCCCTCTGATAGACTCTGCGGGGTCTCATGTCGACACACCGGCAAGAT 932
QY 136 -----TrpGlnLeuHisArg-----ValLysTrpTrpGlnAlaMetAla----- 148
DB 933 TAAAGACTTTGAAGATACACAGAAATCTTCCACGTACTGGGAAGGCAAGACACATGGA 992
QY 149 -----LeuValTyrHisAlaThr-----GlyAspGluLysTyrAlaArgGluTrpVa 164
DB 993 GACGTCACGCGGATCTATGGCAATTCCTCCCTGACCCCAAGCTACTGAAAAGATGGGA 1052
QY 164 lTyrGlnTyrSerAspTrpAlaArgLysAsnProLeuGlyLeuSerGlnAspAsnAspLys 184
DB 1053 GAAGTTCACAGAGGAGCCAAAACCGAGACACACAGGAAATCGGAGGAGCCACAGAAT 1112
QY 184 sPheValTrpArgProLeuGluValSerAspArgValGlnSerLeuProProThrPheSe 204
DB 1113 ATATTTCTTC-----CATGAACCTTAGCCCTGGAAGTTG 1145
QY 204 rLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGluPheLeuAsnSerTy 224
DB 1146 CTTTTCCTGCCCCAAGAGGACCTACATATTATATACACTAATGGAATTTATCAGGAGT-- 1203
QY 224 rHisGlnGlnAlaAspTyrLeuSerThrHisTyrAlaGluGlnGlyAsnHisArgLeuPh 244
DB 1204 -----GAATATAGAAAGAGGGTTCAGAGGGTCGTCACCTCCCAACACTTT 1250
QY 244 eGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPheLysAspSerProAr 264
DB 1251 CAATAGCGGCTCTGGATCACCTCTGGC-----CA 1280
QY 264 gTrpArgGlnThrGlyIleSerValLeuAsnThrGluIleLysGlnValTyrAlaAs 284
DB 1281 CTGGCAGCACTACAGGAGAACATGTTCTCTTTGAGGTGGAGAGGAG----- 1329
QY 284 pGlyMetGlnPheGluLeuSerProIleTyr-----HisValAlaAlaIleAsp-- 300
DB 1330 -----CAGTTTGCACTTAAGCCCATGAATCCCGGCAGACCTGCTGATGTTGATCA 1382
QY 301 -----lPheLeuLysAlaTyrGlySerAlaLysAr 311
DB 1383 CCGGCCACCGTCTCTGGCGAGAGCTGCCCTCGCGCTAGCTTTGGTGTGTCATAG 1442
311 gValAsnLeuGlu-----LysGluPheProGlnSer-- 321
1443 GAATGAGCTCTCGGGGGGCTCTCACCGGGCTCACCGGGTCCGAAGATTCACGAGATGA 1502
322 -----TyrValGlnThrValGluAsnMetIleMetAlaLeuIleSerI1 336
1503 CGCACACATCTCTGTGCCATGGAGACATGAAGATGAAATCAAAAGGTGTTTGGATTT 1562
336 eSerLeuProAspTyrAsnThrProMetPheGlyAspSerTrpIleThrAspLysAsnPh 356
1563 TCTTCGCAAGTATATAGT-----GTCTTTGGATTTTCATTTAAATGAATCTTCTAC 1616
356 eArgMetAlaGlnPheAlaSerTrpAlaArgValPheProAlaAsnGlnAlaIleLysTy 376
1617 TCGCCCAAGAAATTCCTTGAGATATTGAATATGG-----AACCAAGCT----- 1662
376 rPheAlaThrAspGlyLysGlnGlyLysAlaProAsnPheLeuSerLysAlaLeuSerAs 396
1663 -----GAGAAACAACCTTGAAAA 1679
396 nAlaGlyPheTyrThrPheArgSerGlyTrpAspLysAsnAlaThrValMetValLeuLy 416
1680 CAGT---TTGAATGAGTTTCGTGAGAGTGGGAACATAAT----- 1716
416 sAlaSerProProGlyGluPheHisAlaGlnProAspAsnGlyThrPheGluLeuPheI1 436
1717 -----CCTGGAGATGGAGCCTTC----- 1734
436 eLysGlyArgAsnPheThrProAspAlaGlyValPheValTyrSerGlyAspGluAlaI1 456
1735 -----TATGGCCCAAGATTCACATACAGATA-----AAAGATGCCAT 1772
456 eMetLysLeuArgAsnTrpTyrArgGlnThrArgIleHisSerThrLeuThrLeuAspAs 476
1773 CGTGC-----TACCACCACTGT-----GCAACCATCCAGCTGAGCTT 1811
476 nGlnAsnMetValIleThrLysAlaArgGluAsnLysTrpGluThrGlyAsnAsnLeuAs 496
1812 TCAGTTGCCATC-----AGTTTAAAC----- 1833
496 pValLeuThrTyrThrAsnProSerTyrProAsnLeuAspHisGlnArgSerValLeuPh 516
1834 ---CTTACTTACGTC-----AGCCACGATGGCGATGATAAGAAAGGCCAGTGATT-- 1881
516 eLleAsnLysLysTyrPheLeuValIleAspArgAlaIleGlyGluAlaThrGlyAsnLe 536
1882 -GTTCCACAGAGCATCTCGGGTCGGTGGAAAGATGATTGCCATCTCCACAGAAACTA 1940
536 uGlyValHisTrpGln-----LeuLysGluAspSerAsnPr 548
1941 TGGGGCAATGGCCTTTCTGCTCTCTCCGAGGTGATGGTCCCGAGTGGGACC 2000
548 oValPheAspLysThrLysAsnArgValTyrThrTyrArgAspGlyAsnAsnLeuMe 568
2001 GACATGTGATGAATATGCCAAAGGTACGGCAACAATTCATGATGCTAAATTCATGCG 2060
568 tIleGlnSerLeuAsnAlaAspArgThrSerLeuAsnGluGluGluGlyLysValSerTy 588
2061 GCACACTGACCTGGATCCAGGCTGTACC---TTGAATAAGAG-- 2100
588 rValTyrAsnLysGluLeuLysArgProAlaPheValPhe-----GluLysProLy 605
2101 -ATCAGAAACGACAGTTAGCACATATAACTTCATCTTGTGTGTGTGTAAGAAA--GA 2156
605 sLysAsnAlaGlyThrGlnAsnPheValSerIleValTyrProTyrAspGlyGlnLysAl 625
2157 GAAAGCCAGTCGACCGTGAAC----- 2178
625 aProGluIleSerIleArgGluAsnLysGlyAsnAspPheGluLysGlyLysLeuAsnLe 645
2179 -----ATCCGACGAGAGACAAATAGGTCAC-----GGAGACGAGACAGT 2219
645 uThrLeuThrIleAsnGlyLysGlnLeu 655

```

```

Db      2220 GGAGGAGCGGTGGCGCGCTGCAGCAGCTC 2250
RESULT 3
Bg566122
LOCUS   957 bp mRNA linear EST 10-APR-2001
DEFINITION 602582746F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4710659 5',
mRNA sequence.
ACCESSION Bg566122
VERSION   Bg566122.1 GI:13573775
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 957)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM1549 row: k column: 12
High quality sequence stop: 739.
Location/Qualifiers
1. 957
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4710659"
/clone_lib="NIH_MGC_76"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site 1:
SfiI (ggccattatgcc); Site 2: SfiI (ggccattatgcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGGCCATTATGCC-3', and 3' adaptor sequence:
5'-ATTCTAGGCGGAGCGCGGACATGTT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."
BASE COUNT 269 a 256 c 254 g 178 t
ORIGIN

Alignment Scores:
Pred. No.: 0.294 Length: 957
Score: 99.50 Matches: 64
Percent Similarity: 33.56% Conservative: 33
Best Local Similarity: 22.15% Mismatches: 121
Query Match: 2.85% Indels: 71
DB: 12 Gaps: 13

US-09-802-285A-2 (1-659) x Bg566122 (1-957)

Qy      187 TrpArgProLeuGluValSerAspArgValGlnSerLeuProThrPheSerLeuPhe 206
Db      185 TTTTGAACAGGAGTAGTAAATAGTCGGTTCAGGAGATCCAAAGCAACTCTTTTACTTC 244
Qy      207 ValAsnSerProAlaPheThrProAlaPheLeuMetGluPheLeuAsnSerTyrHisGln 226
Db      245 ACCCCCAACAGACAGGACACACATTTCTTCAGAGAGTAC----CAGACCCGACAGAC 301
Qy      227 GlnAlaAspTyrLeuSerThrHis-----TyrAlaGlnGlnGlyAsnHisArgLeu 243
Db      302 CAGTGCATCTATAACACACCTACCTGAAATGCCAGCGGAAATGGACCATCTCCAGA 361
Qy      244 PheGluAlaGlnArgAsnLeuPheAlaGly----- 253

```

```

Db      362 TACGTGGAGCCAGAGCAATTCGCTCACTTGTGATCCTCAGGAGACCAAGACCTAC 421
Qy      254 ---ValSerPheProGluPheLysAspSerProArgTrpArgGlnThrGlyIleSerVal 272
Db      422 ATGCTTGTCTTTT---GAGCTGAACGATGAGAAGAACTGG-----GGGCTGTCT--- 466
Qy      273 LeuAsnThrGluIleLysGlnValTyrAlaAsp----- 284
Db      467 -----GTCTATCTGACAAAGCCAGACGACGACCAAGAGCAAA 502
Qy      285 ---GlyMetGlnPheGluLeuSerProIleTyrHisValAlaAlaIleAspIlePheLeu 303
Db      503 CTGGGAGAGTTCTACGAGCTCTCGACTGTGTCGTCATTCGCCAAGTCAGATGTGCTGT 562
Qy      304 LysAlaTyrGlySerAlaLysArgValAsn---LeuGluLysGluPheProGlnSerTyr 322
Db      563 ACACCGATTGGAAAAACGGAATAGTGTGAGCCACTGGAGAAAGCAGACGACGAGGAGAG 622
Qy      323 ValGlnThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuProAspTyrAsn 342
Db      623 GAAACCAAGAGGAGGAGGAGTCTAGCCAGGACA----- 658
Qy      343 ThrProMetPheGlyAspSerTrpIleThrAspLysAsnPheArgMetAlaGlnPheAla 362
Db      659 -----CAGCCTTGGATCCAGACAGAGACTTGGGGGCCATCTGACCCCTC 703
Qy      363 SerTrpAlaArgVal-----PheProAlaAsn-----GlnAlaIle 374
Db      704 CAAACCCGACAAATGTTGTACTCTCAGCTTTTCCCTTGGCATCCAGTAGGAGCTATC 763
Qy      375 LysTyrPheAlaThrAspGlyLysGlnGlyLysAlaProAsnPheLeuSerLysAlaLeu 394
Db      764 CTGTGGTTTGGACCAAGAAACAAAGGCTTCAACACACAAAGAAATAATGTTTA 823
Qy      395 SerAsnAlaGlyPheTyrThrPheArgSerGlyTrpAspLysAsnAlaThrValMetVal 414
Db      824 AGCGGCGCT-----TGGGACGCGAGAAAGAGGCTCTTAGTA 859
Qy      415 Leu-----LysAlaSerProGlyGluPheHisAlaGlnProAspAsnGly 430
Db      860 CCACACAGGTGGGACCAACGCGCGCCCAAAACATGACAGCAGCAGGAGGAGGAGGCGC 919
Qy      431 ThrPheGluLeuPheIleLysGlyArg 439
Db      920 ACCAAAGAGAGCCGGTGAAGGAACGG 946

RESULT 4
AW942324/c
LOCUS   580 bp mRNA linear EST 23-APR-2001
DEFINITION LD14783.3prime LD Drosophila melanogaster embryo BlueScript
Drosophila melanogaster cDNA clone LD14783 3, mRNA sequence.
ACCESSION AW942324
VERSION   AW942324.1 GI:8120024
KEYWORDS  EST.
SOURCE    fruit fly.
ORGANISM  Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 580)
Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,
Lewis, S. and Rubin, G.M.
BDGP/HMHI Drosophila EST Project
Unpublished (2001)
Other ESTs: LD14783.5prime
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Based upon the presence of a XhoI site followed by a run of 14 or

```



```

Db 420 TTAGATCAAGCAATAGTGGTGTATATCTTTTCAATCAAACTTTTCCCAATTGAT 479
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Qy 403 ArgSerGlyTTPAspLysAsnAlaThrValMetValLeuLysAlaSerProGlyGlu 422
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 480 AACATGGGGTACGATGTAGATCCATCATTTAGAAATTTATAACAAAGT-----GGTAAC 533
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Qy 423 PheHisAlaGlnProAspAsnGlyThrPheGluLeuPheIleLys--GlyArgAsnPhe 441
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 534 TATCAT-----AATTTCCATTTTGTGTAAAGATGAATAGTCGTTTT 575
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Qy 442 ThrProAspAlaGly---ValPheValTyrSerGlyAspGlu 454
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 576 ACTTATAATGGAGGTGAAGTGTTTTAATTTAAAGGTGATGAT 617
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

RESULT 6
BJ424970          655 bp mRNA linear EST 11-MAR-2002
LOCUS             Dictyostelium discoideum cDNA library, VF Dictyostelium
DEFINITION        Dictyostelium cDNA clone ddv54002 5', mRNA sequence.
ACCESSION         BJ424970
VERSION           BJ424970.1 GI:19341677
KEYWORDS          EST.
SOURCE            Dictyostelium discoideum.
ORGANISM          Dictyostelium discoideum
REFERENCE         1 (bases 1 to 655)
AUTHORS           Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE             Full length cDNA of Dictyostelium discoideum at the vegetative
                  stage
JOURNAL            Unpublished (2002)
COMMENT           Contact: Tadasu Shin-i
                  Center For Genetic Resource Information
                  National Institute of Genetics
                  1111 Yata, Mishima, Shizuoka 411-8540, Japan
                  Tel: 81-559-81-6856
                  Fax: 81-559-81-6855
                  Email: tshini@genes.nig.ac.jp.
FEATURES          source
                  1..655
                   /organism="Dictyostelium discoideum"
                   /strain="AX4"
                   /db_xref="taxon:44689"
                   /clone="ddv54002"
                   /clone_lib="Dictyostelium discoideum cDNA library, VF"
                   /sex="mat A"
                   /dev_stage="Growth phase"
BASE COUNT        250 a 94 c 88 g 223 t
ORIGIN

Alignment Scores:
Pred. No.:         0.289          Length:         655
Score:             97.50          Matches:         49
Percent Similarity: 36.9%          Conservative:    30
Best local Similarity: 22.90%      Mismatches:     70
Query Match:       2.79%          Indels:         65
DB:                13            Gaps:         12

US-09-802-285A-2 (1-659) x BJ424970 (1-655)

Qy 269 GlyIleSerValLeuAsnThrGluIleLysLys-----GlnValTyrAla 283
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 99 GGTGTGAGTTTAATGATACATCAATCAATTAATCAGTTCCAATGGCCCAACAATTATATT 158
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Qy 284 AspGlyMetGlnPheGluLeuSerProIleTyrHisValAlaAlaIleAspIlePheLeu 303
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 159 GAAGGTACAAATTTTGTCAATGGCCAGAATATAAT-----194
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Qy 304 LysAlaTyrGlySerAlaLysArgValAsnLeuGluLysGluPheProGlnSerTyrVal 323
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 195 -----TATAATTTTGAAATAACCAACCCAGGAGTAAATGTT 230
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Qy 324 GlnThrValGluAsnMetIleMetAlaLeuIle-----SerIleSerLeuProAspTyr 341
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

```

```

Db 231 AAAACCCCTT---GGATGTTAAAGCAAAATTTGAATCAACAAACAAAGTTCCAGAACTT 287
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Qy 342 AsnThrProMetPheGlyAspSerTrpIleThrAspLysAsnPheArgMetAlaGlnPhe 361
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 288 TCTTCA-----GATGATACTAAAACTTCCACCAATACATTAGGTACTATGCAATAT 338
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Qy 362 AlaSerTrpAlaArgValPheProAlaAsnGlnAlaIleLysTyrPheAlaThrAspGly 381
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 339 CCA-----GAGTTATTTAATATCACTTTACATCAAAAT 371
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Qy 382 LysGlnGlyLysAlaProAsnPheLeuSerLysAlaLeu-----394
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 372 GTTGATGCACCCAGCATCCAAAACTCTGTGTAATAAATTTTACCAATGAAAAATACAATTCAA 431
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Qy 395 -----SerAsnAlaGlyPheTyrThrPhe-----402
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 432 TTAGATCAAGCAAAATAGTGGTGTATATCTTTTCAATCAAACTTTTCCCAATTGAT 491
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Qy 403 ArgSerGlyTTPAspLysAsnAlaThrValMetValLeuLysAlaSerProGlyGlu 422
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 492 AACATGGGCTAGATGTAGATCCATCATTTAGAAATTTATAACAAAGT-----GGTAAC 545
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Qy 423 PheHisAlaGlnProAspAsnGlyThrPheGluLeuPheIleLys--GlyArgAsnPhe 441
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 546 TATCAT-----AATTTCCATTTTGTGTAAGATGAATAGTCGTTTT 587
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Qy 442 ThrProAspAlaGly---ValPheValTyrSerGlyAspGlu 454
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 588 ACTTATAATGGGGTGAAGTGTTTTAATTTAAAGGTGATGAT 629
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

RESULT 7
BJ421576          659 bp mRNA linear EST 10-MAR-2002
LOCUS             Dictyostelium discoideum cDNA library, VF Dictyostelium
DEFINITION        Dictyostelium cDNA clone ddv43k03 5', mRNA sequence.
ACCESSION         BJ421576
VERSION           BJ421576.1 GI:19333816
KEYWORDS          EST.
SOURCE            Dictyostelium discoideum.
ORGANISM          Dictyostelium discoideum
REFERENCE         1 (bases 1 to 659)
AUTHORS           Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE             Full length cDNA of Dictyostelium discoideum at the vegetative
                  stage
JOURNAL            Unpublished (2002)
COMMENT           Contact: Tadasu Shin-i
                  Center For Genetic Resource Information
                  National Institute of Genetics
                  1111 Yata, Mishima, Shizuoka 411-8540, Japan
                  Tel: 81-559-81-6856
                  Fax: 81-559-81-6855
                  Email: tshini@genes.nig.ac.jp.
FEATURES          source
                  1..659
                   /organism="Dictyostelium discoideum"
                   /strain="AX4"
                   /db_xref="taxon:44689"
                   /clone="ddv43k03"
                   /clone_lib="Dictyostelium discoideum cDNA library, VF"
                   /sex="mat A"
                   /dev_stage="Growth phase"
BASE COUNT        254 a 94 c 88 g 223 t
ORIGIN

Alignment Scores:
Pred. No.:         0.292          Length:         659
Score:             97.50          Matches:         49
Percent Similarity: 36.9%          Conservative:    30
Best local Similarity: 22.90%      Mismatches:     70
Query Match:       2.79%          Indels:         65
DB:                13            Gaps:         12

```



```

QY 173 LysAsnProLeuGlyLeuSerGlnAspAsnAspLysPheValTyrArgProLeuGluVal 192
Db 290 AAG-----GATACGATCTCTTTCATCTGG-----313
QY 193 SerAspArgValGlnSerLeuProThrPheSerLeuPheValAsnSerProAlaPhe 212
Db 314 -----ACAGTCCAG-----TTTGTGAGCTGACATCCAGGCCATCTCTC 352
QY 213 ThrProAlaPheLeuMetGluPheLeuAsnSerTyrHisGlnGlnAlaAspTyrLeuSer 232
Db 333 TGGAAC-----CTACAGCCCTATTCCACCGTGTACCAATAGAAATATTTTCTGAGC 406
QY 233 ThrHisTyrAlaGluGlnGlyAsnHisArgLeuPheGluAlaGlnArgAsnLeuPheAla 252
Db 407 CCCAAGTACTCGAGCAGTATCCCAATGACATA-----GCCCTGCTGAGTGCATCT 460
QY 253 GlyValSerPheProGluPheLysAspSerProArgTyrArgGlnThrGlyIleSerVal 272
Db 461 CCAGTCACCTACAATAAATCTTCATCCAGGCC-----ATCTGCCTC 499
QY 273 LeuAsnThrGluIleLysLysGlnValTyrAlaAspGlyMetGlnPheGluLeuSerPro 292
Db 500 CTGAATCCAGTACAG-----TTTGAG-----523
QY 293 IleTyrHisValAlaAlaIleAspIlePheLeuLysAlaTyrGlySerAlaLysArgVal 312
Db 524 -----AACCGAACTGACTGCTGGGTGACCGGCTGGGGGCT-----ATT 562
QY 313 AsnLeuGluLysGluPheProGlnSerTyrValGlnThrValGluAsnMetIleMetAla 332
Db 563 GGAGAAAGTATGAGTGTCTGCCATCTCCC-----AACACTCTCCAGGAAGTCAGGTAGCT 616
QY 333 LeuIleSerIleSerLeu-----ProAspTyrAsnThrProMet 345
Db 617 ATTATCAACACAGCATGTGTAAACCATATGTACAAAAAGCCAGACTTCCTCCGACGAACATC 676
QY 346 PheGlyAsp 348
Db 677 TGGGGAGAC 685

RESULT 11
BO744898 671 bp mRNA linear EST 17-JUL-2002
LOCUS 946112B02.y1 946 - tassel primordium prepared by Schmidt lab Zea
DEFINITION mays cDNA, mRNA sequence.
ACCESSION BO744898
VERSION BO744898.1 GI:21891685
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.

REFERENCE
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
JOURNAL University
COMMENT Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 946112 row: B column: 02.
Location/Qualifiers
1..671
/organism="Zea mays"
/cultivar="OH43"
/db_xref="taxon:4577"
/clone_lib="946 - tassel primordium prepared by Schmidt"

FEATURES
source

```

```

lab"
/tissue type="tassels"
/dev_stage="just after the transition from vegetative to
inflorescence development"
/lab host="XL0LR"
/Note="Organ: tassels; Vector: HybridZAP; Site 1: EcoRI;
Site 2: XhoI; George Chuck dissected immature tassels
between 1mm and 3mm. Sharon Stanfield prepared the cDNA
library in HybridZAP. Sample insert size range was 350 bp
to 3 kb with a 1 kb average."
BASE COUNT 181 a 169 c 165 g 156 t
ORIGIN

Alignment Scores:
Pred. No.: 0.62 Length: 671
Score: 95.00 Matches: 46
Percent Similarity: 36.04% Conservatve: 25
Best local Similarity: 23.35% Mismatches: 64
Query Match: 2.72% Indels: 62
DB: 14 Gaps: 9

US-09-802-285a-2 (1-659) x BO744898 (1-671)
QY 7 LysArgIleIleValPheAlaValIleAlaLeuSerSerGlyAsnIleLeuAlaGlnSer 26
Db 116 AAGCGATTACAGATTTCTTATTGCTGCACATCTCTCTGACAATGTCTG-----TAT 169
QY 27 SerSerIleThrArgLysAspPheAspHis-----IleAsnLeu 39
Db 170 ATCCAGTCCGTGTATCTCTGATCTGGACCAACAATTGTTGGGAAAGGCCAACAATGACC 229
QY 40 GluTyrSerGlyLeuGluLysValAsnLysAlaValAlaAlaGlyAsnTyrAspAla 59
Db 230 GAGAAACACCTCTCACACATCAACAAGAGACTCCCGATCAGATGTTCTCTCTGAG 289
QY 60 AlaLysAlaLeuLeuAlaTyrTyrArgGluLysSerLysAlaArgGluProAspPheSer 79
Db 290 GCAGCAGCAGCATCGGAGCTGCTTCAATGGTCTTCAATCCAGAGATACCACTATTTCT 349
QY 80 AsnAlaGluLysProAlaAspIleArgGlnProIleAspLysValThrArgGluMetAla 99
Db 349 -----349
QY 100 AspLysAlaLeuValHisGlnPheGlnProHisLysGlyTyrGlyTyrPheAspTyrGly 119
Db 350 GATGTGCTTCTTCAGCAT-----GCTCAGAACTATTTCACCTTTTGCTGACACTTAT 400
QY 120 LysAspIleAsnTrpGlnMetTrpPro---ValLysAsp-----131
Db 401 AGAGGCTCTCAAGCGAGAGCTATCCCAAGCTCCAGGATTTCTATATACTCCACCACTTAC 460
QY 132 ---AsnGluValAlaGTrpGlnLeuHisArgValLysTrpTrpGlnAlaMetAlaLeuVal 150
Db 461 GTTGATGAGCTTCTATGG-----GCAGCGAGTTGGCTC 493
QY 151 TyrHisAlaThrGlyAspGluLysTyrAlaArgGluTrpValTyrGln-----166
Db 494 TACCAGCGCCACAGCGGACCGACGATCTCTCAGCTATGTAACTGTACAGACGGGAAAGT 553
QY 167 TyrSerAspTrpAlaArg-----LysAsnPro 175
Db 554 TATGCTGATGGGAAGGCCCAACATGTTGATGGGATGACAAAAATCCA 604

RESULT 12
CNS06JJZ 1001 bp DNA linear GSS 17-JUN-2001
LOCUS T7 end of clone AS0AA029D05 of library AS0AA from strain CLIB 533
DEFINITION of Saccharomyces bayanus, genomic survey sequence.
ACCESSION AL401685
VERSION AL401685.1 GI:12159698
KEYWORDS GSS.
SOURCE Saccharomyces bayanus.
ORGANISM Saccharomyces bayanus.

```


Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

REFERENCE AUTHORS

Souci t,J.B., Aigle,M., Artiguenave,F., Blandin,G.,
Boltin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
Malpertuy,A., Neuveglise,C., Olier-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvet,M.,
Wincker,P. and Weissenbach,J.

Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies

FEBS Lett. 487 (1), 3-12 (2000)

11152876

REFERENCE AUTHORS

Bon,E., Neuveglise,C., Casaregola,S., Artiguenave,F., Wincker,P.,
Aigle,M. and Durrens,P.

Genomic exploration of the hemiascomycetous yeasts: 5.

Saccharomycetes bayanus var. uvarum

FEBS Lett. 487 (1), 37-41 (2000)

20584715

11152880

REFERENCE AUTHORS

Direct Submission
Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomycetes bayanus var. uvarum, Saccharomycetes
exiguus, Saccharomycetes servazii, Zygosaccharomycetes rouxii,
Saccharomycetes kluyveri, Kluyveromycetes thermotolerans, Kluyveromycetes
lactis var. lactis, Kluyveromycetes marxianus var. marxianus, Pichia
angusta, Debaryomyces Hansenii var. Hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.

FEATURES source

1. .1001
/organism="Saccharomycetes bayanus"
/strain="CLIB 533"
/variety="uvarum"
/db_xref="taxon:4931"
/clone="AS0AA029D05"
/clone_lib="AS0AA"
/note="end : 17"
<3. .>1001
/note="similar to Saccharomycetes cerevisiae ORF YNL262w [
POL2 : DNA-directed DNA polymerase epsilon, catalytic
subunit A]"
/evidence=not_experimental

misc_feature

BASE COUNT 350 a 161 c 209 g 280 t 1 others
ORIGIN

Alignment Scores:

Pred. No.: 1.55 Length: 1001
Score: 94.00 Matches: 69
Percent Similarity: 35.45% Conservative: 48
Best Local Similarity: 20.91% Mismatches: 117
Query Match: 2.69% Indels: 96
DB: 17 Gaps: 19

US-09-802-285A-2 (1-659) x CNS06JUZ (1-1001)

QY 336 IleSerLeuProAspTyrAsnThrPrometPheGlyAspSerTirpileThrAsp----- 353

Db 108 TTGGATTCCCTGAATTAATAATCCAGATATCATGATAATGCGTACTAGTGTACT 167

QY 354 ----LysAsnPhaGlyMetAlaGlnPheAlaSerTirPalaArgValPheProLa----- 370

Db 168 GTTGATAATCTACACATAACACAAATATTGGCATCTGCAATGATCAATGATGAGAGGT 227

QY 371 -----AsnGlnAlaIleLysTyrPheAlaThrAspGlyLysGlnGlyLysAla 386
Db 228 AOTGATCTGGTTAATAGCGGTATGGTATGGAGGATAAAAAATGCCAACAGCACTCTACG 287
QY 387 ProAsnPhaSerLysAlaLeuSerAsnAlaGlyPheTyrThrPheArg----- 403
Db 288 TCGGATTTTGTCAGGATGCCCTTTCTAATGATGCTTTAAATGCTCCTAAGGGCATGCTA 347
QY 404 ----SerGlyTirAspLysAsnAlaThrValMetValLeuLysAlaSerProProGlyGlu 422
Db 348 AAGGAATGCTGGATGAA----- 365
QY 423 PheHisAlaGlnProAspAsnGlyThrPheGluLeuPheIleLysGly----- 438
Db 366 -----GCTTGAAGAGATAATGCACTGCGGATTTTGTGTAACACTCGTTAGCAAGCTGG 419
QY 439 ---ArgAsnPhaThrProAspAlaGlyValPhe-----ValTyr 450
Db 420 GTCCAAAT-----CCTAATGCAAAATATTATTGATGGTTTCTGAGATATCATGTCCAT 473
QY 451 Ser---GlyAspGluAlaIleMetLysLeuArgAsnTirTyrArgGlnThrArgIleHis 469
Db 474 AACTTGACCAAAAAATCACTACTCCAATTAGTGAATGATTT-----AGCACTCTGGGC 527
QY 470 SerThrLeuThrLeuAspAsnGlnAsnMetValIleThrLysAlaArgGlnAsnLysTirP 489
Db 528 TCACTATCTGTATGTCAGATAGAACCAAAATTTTAAATCAAAACA-----ACAAGTAC 581
QY 490 GluThrGlyAsnAsnLeuAspValLeuThrTyr-----ThrAsnPro 503
Db 582 TCGCCTGAAAATTTGTTACGCTTACAGTCAATATATCATGAAACCGCTTAGGACAAATCCA 641
QY 504 SerTyrProAsnLeuAspHisGlnArgSerValLeuPheIleAsnLysLysTyrPheLeu 523
Db 642 ATGTTTCAGTTATTGGAT-----TTAAATATTAAAGCGTTATTGGATCTG 686
QY 524 ValIle-----AspArg-----AlaIleGlyGluAlaThrGly----- 534
Db 687 TTAATATGATGATGATAAAATTCAACTTCAGTGGTTAGCTTTGTGGAAGATAGAGGAAAA 746
QY 535 -----AsnLeuGlyValHisTirPginLeuLysGluAspSerAsnProVal 549
Db 747 GAAAGCAAGACTATAACGCAATTTTCGGAATGGCACTAAAGAAATTTTTCACCAATA 806
QY 550 PheAsp-----LysThrLys 554
Db 807 TATCAACCTGAATTTGAGGATTTGATGATGATCATCTTAGACAGTATGTTAAGACCAAG 866
QY 555 AsnArgValTyrThrTyrArgAspGlyAsnAsnLeuMetIleGlnSerLeuAsnAla 574
Db 867 CAGAAC---TATTTACATCAATTCAGGTACACAAAGACCTACCCAAATACTCAACACAG 923
QY 575 AspArgThrSerLeuAsnGluGlyLysValSerTyrValTyrAsnLysGluLeu 594
Db 924 AAGAAAGAA-----GACAAAGAGATGCG-----ATTGACAACTCACTG 962
QY 595 LysArgProAlaPheValPheGluLysPro 604
Db 963 AATGGGGTCYCACACATTTTCTCCAAGCCA 992
RESULT 13
BM413117
LOCUS
DEFINITION
EST587444 tomato breaker fruit Lycopersicon esculentum cDNA clone
C1EG62E22 5' end, mRNA sequence.
ACCESSION
BM413117
VERSION
BM413117.1 GI:18264747
KEYWORDS
EST.
SOURCE
tomato.
ORGANISM
Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Lycopersicon.
1 (bases 1 to 756)
Alcala,J., Vrebalov,J., White,R., Vision,T., Karamycheva,S.A., Tsai
,J., Bougri,O., Kirkness,E., Otterback,T., Van Aken,S., Ronning
,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato fruit tissue, breaker stage (2002)
Unpublished (2002)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
This clone is available through the Clemson University Genomics
Institute
Seq primer: T3
Location/Qualifiers
1..756
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLSG62E22"
/clone_lib="tomato breaker fruit"
/cissue_type="pericarp"
/dev_stage="breaker"
/lab_host="SOLR"
/note="Vector: pBluescriptSKmCUadapt; Site 1: EcoRI;
Site 2: XhoI; supplier: Boyce Thompson Institute; Fruit
sequencing: The Institute for Genomic Research. Fruit
were harvested at the breaker stage (first sign of
lycopene accumulation on the blossom end of fruit). Fruit
were cut in half and the seeds and locules were discarded
prior to freezing the pericarp."
BASE COUNT 267 a 105 c 172 g 212 t
ORIGIN

Alignment Scores:
Pred. No.: 1.33 Length: 756
Score: 93.00 Matches: 49
Percent Similarity: 36.50% Conservative: 24
Best Local Similarity: 24.50% Mismatches: 72
Query Match: 2.66% Indels: 55
DB: 13 Gaps: 11

US-09-802-285a-2 (1-659) x BM413117 (1-756)

QY 378 AlathrAspGlyLysGlnGlyLysAlaProAsnPheLeuSerLysAlaLeuSerAsnAla 397
DB 233 AGTGTCAATGGAAGAAAGGGTTACACCGGAAGAGTGTAAAGCTT-----GGTGGT 286
QY 398 GlyPheTyrThrArgSerGlyTrpAspLysAsnAlaThrValMetValLeuLysAla 417
DB 287 GGTACACACATTC-----CTACAACT 310
QY 418 SerProGlyGluPheHisAla---GlnProAspAsnGlyThrPheGlu-----433
DB 311 TCATTGCCAGAAATCTAAGAGTCTATACTCTGAGGATGAACATTTGAATCATCAACAA 370
QY 434 -----LeuPheLeuLysGlyArgAsnPheThrProAspAlaGlyValPhe 448
DB 371 AATGTATTAGATCAATATTGTAGCTGGA-----TTGCGCATCGAAGTC---CTTCAT 421
QY 449 ValTyrSerGly---AspGluAlaIleMetLysLeuArgAsnTrp---TyrArg-GlnTh 466
DB 422 GTTTATTACAGCACCAAGAAATGTGTACAGTTTAGCATTTGGGTTACATGATGGT 481
QY 466 rArgIleHisSerThrLeuThrLeuAspAsnGlnAsnMetValIleThrLysAlaArgG 486
DB 482 CCATTAAAGGATATGTTGCACTGGACAACTTGTAGACTCTTTTGAATTGGCATTTT 541
QY 486 nAsnLysTrpGluThrGlyAsnLeuAspValLeuThrTyrThrAsnProSerTyrPr 506
DB 542 GAAT---TGG-----548

QY 506 oAsnLeuAspHisGlnArgSerValLeuPheIleAsnLysLysTyrPheLeuValIleAs 526
DB 549 -----AGAAAGACAGTAACACAGATTGTTAAGGACAGAGTCTTTTTCATAGAGA 598
QY 526 pArgAlaIleGly-----GluAlaThrGlyAsnLeuGlyValHisTrpG 541
DB 599 GAATTGCTGGCCCACTTATGAAAGGAGGAAAAAATGTTGAATCCACAAAGTGAGATGCA 658
QY 541 nLeuLysGluAspSerAsnProValPheAspLysThrLysAsnArgValTyrThr 560
DB 659 TTTATTGAGGATTCAAAATCCCTTTCATGAAATAAATAATAATCTACTAGCAACA 716
RESULT 14
BMI65623 762 bp mRNA linear EST 04-DEC-2001
LOCUS EST568146 PyBS Plasmodium yoelii yoe111 cDNA clone PYCMP96 5' end,
DEFINITION mRNA sequence.
ACCESSION BMI65623
VERSION BMI65623.1 GI:17298855
KEYWORDS EST.
SOURCE Plasmodium yoelii yoe111.
ORGANISM Plasmodium yoelii yoe111.
REFERENCE 1 (bases 1 to 762)
AUTHORS Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
1 (Bases 1 to 762)
Fraser,C.M. and Carucci,D.J.
Plasmodium yoelii EST project at TIGR
Unpublished (2001)
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ATCC
<http://www.malaria.mr4.org/mr4pages/index.html>
Seq primer: ADF.
Location/Qualifiers
1..762
/organism="Plasmodium yoelii yoe111"
/strain="17XL"
/db_xref="taxon:73239"
/clone="PYCMP96"
/clone_lib="PyBS"
/dev_stage="Asexual blood stages"
/note="Vector: pAD-GAL4; At 20-25% parasitemia, blood was
collected from BALB/cBYJ mice infected with Py17XL
parasites, and leukocytes removed by passage over
microcrystalline cellulose columns. Total RNA was
isolated using the guanidium isothiocyanate method, and
mRNA isolated using oligo(dT)-cellulose chromatography.
First strand cDNA synthesis was completed using a 50-base
primer and reverse transcriptase in the presence of
5-methyl dCTP. After second strand synthesis, uneven
termini were treated with Pfu DNA polymerase and EcoRI
adaptors ligated to the blunt ends. The sample was cleaved
with XhoI and separated on a Sephacryl S-500 column.
Size-fractionated cDNA was precipitated and ligated to
HybriZAP arms directionally using EcoRI-XhoI cleaved arms.
After packaging, the phagemid vector (pAD-GAL4) was
excised from the HybriZAP vector and plasmid DNA
isolated."
BASE COUNT 368 a 81 c 118 g 195 t
ORIGIN

Alignment Scores:
Pred. No.: 1.35 Length: 762
Score: 93.00 Matches: 44
Percent Similarity: 38.95% Conservative: 30
Best Local Similarity: 23.16% Mismatches: 56

```
Query Match: 2.66% Indels: 60
DB: 13 Gaps: 11
US-09-802-285A-2 (1-659) x BML65623 (1-762)
QY 487 AsnLysTrpGluThrGlyAsnAsnLeuAspValLeuThrTyrThrAsnProSerTyrPro 506
Dd 17 AACAAAGACACACCAATATGATCGATGACAGATTAATATATGAGAAAAGTG 76
QY 507 AsnLeuAspHisGlnArgSerVal-----LeuPheIleAsnLys 519
Dd 77 AATGAATATACATAAAATATATAGATTCTCAAGTGATCAAGCCCAATTAAGAAATAAG 136
QY 520 -----LysTyrPheLeuValIleAspArgAlaIleGlyGluAlaThrGlyAsnLeuGly 537
Dd 137 ATAACTAAATATAGTAACATATTGAT----- 163
QY 538 ValHisTrpGlnLeuLysGluAspSerAsn-----ProValPheAspLysThrLys 554
Dd 164 -----GATCTTAAAGAGAGATAATGAATTGTGGAAGTTATTAGCACATTAAAG 214
QY 555 AsnArgValTyrThrTyrArgAspGlyAsnLeu-----MetIle 569
Dd 215 GAGTATATCAACACATACAAAGGAGACACAGATATTCCACCAATATGACGATCTAATA 274
QY 570 GlnSerLeuAsnAlaAspArgThrSerLeuAsnGluGluGluGlyLysValSer----- 587
Dd 275 AATAAGCTCTAAAGAGATGTCACATTGAATGAAGAAATTCTTAGCGAAGAAAAAAG 334
QY 588 -----TyrValTyrAsnLysGluLeuLysArgProAlaPheValPheGlu 602
Dd 335 AATATAGCTGATGTACTATTACAAAGGAATTGCAA-----ANGTTGCAA 382
QY 603 LysProLysLysAsnAlaGlyThrGlnAsnPheValSerIleValTyrProTyrAsp--- 621
Dd 383 AATGATAAG-----ATGAAAATTATTGAAAAATTGTATATA 418
QY 622 ---GlyGlnLysAlaPro-----GluIleSerIleArgGluAsnLys 634
Dd 419 GCAAATGATAGTGTCTCCCATTTATATAGCCATTAAATGTAACATTACATGAAAAAAA 478
QY 635 GlyAsnAspPheGluLysGlyLysLeuAsn 644
Dd 479 GGAGAT-----GAAAAAATGAAAAAAT 502
RESULT 15
BI647560
LOCUS
DEFINITION
VERSION
ACCESSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM11807 row: m column: 10
High quality sequence stop: 654.
Location/Qualifiers
```

```
source 1. 780
/organism="Mus musculus"
/strain="129,C57BL/6J,FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5318577"
/clone_lib="NCI_CGAP_Mam3"
/tissue_type="tumor, gross tissue"
/dev_stage="10 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: Sali;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
Reference for transgenic model: Xu et al., Nature Genetics
22, 37-43 (1999)."
BASE COUNT 173 a 222 c 248 g 137 t
ORIGIN
Alignment Scores:
Pred. No.: 1.4 Length: 780
Score: 93.00 Matches: 30
Percent Similarity: 41.73% Conservative: 23
Best Local Similarity: 23.62% Mismatches: 34
Query Match: 2.66% Indels: 40
DB: 13 Gaps: 6
US-09-802-285A-2 (1-659) x BI647560 (1-780)
QY 135 ArgTrpGlnLeuHisArgValLysTrpTrpGlnAlaMetAlaLeuValTyrHisAlaThr 154
Dd 261 AGGATGGAGATACACAGATTGCTCTCTGGACATCGATATCTGTG----- 305
QY 155 GlyAspGluLysTyrAlaArgGluTrpValTyrGlnTyrSerAspTrpAlaArgLysAsn 174
Dd 306 ---GGCCATCGATTCCCAAGATCATGGGCTTAGAGAGAGCAGGTTCCACGAGCGGCT 362
QY 175 ProLeuGly----- 177
Dd 363 CCGCTGGTCTCCAGTGTACGTGATGACCACTGGGGGTGACTGTCAGTGGGCTTCCTG 422
QY 178 LeuSerGlnAspAsnAspLysPheValTrpArgProLeuGluValSerAspArgValGln 197
Dd 423 CTTAGCAGCCAGATGATGCTGTCACTCGAGGGGACCAAGAAATGGTATGATCAAG 482
QY 198 SerLeuProPro-----ThrPheSerLeuPheValAsnSer 209
Dd 483 CAGCTTCTCTCCGTGATGTAGACTGGGCGGAGACGTGGACCTAC---CTCATCGTAGGACAC 539
QY 210 ProAlaPheThrProAlaPheLeuMetGluPheLeuAsnSerTyrHisGlnGlnAlaAsp 229
Dd 540 CCGCTCGGACACGT-----CTGACGAACACCTCTCAGTGGTCCGACG 581
QY 230 TyrLeuSer-----ThrHis 234
Dd 582 TACCTGGCTGGCGGCACACAT 602
Search completed: August 4, 2003, 12:30:08
Job time : 2474 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 28, 2003, 20:46:13 ; Search time 109 Seconds
(without alignments)
1560.153 Million cell updates/sec

Title: US-09-802-285A-2

Perfect score: 3494

Sequence: 1 MTTKFKRIIVFAVIALSSG.....KGGKLNLTITNGKQLVLVP 659

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriaph:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3494	100.0	659	2	Q59289 pedobacter
2	250.5	7.2	634	16	Q8E376 streptococc
3	250.5	7.2	634	16	Q8DXF4 streptococc
4	232	6.6	635	16	Q8K866 streptococc
5	231	6.6	635	16	Q9A0T7 streptococc
6	229	6.6	635	16	Q8P1V6 streptococc
7	191.5	5.5	645	2	Q8KN09 streptococc
8	189	5.4	633	16	Q8CZ74 streptococc
9	186.5	5.3	653	16	Q8RJK0 streptococc
10	176	5.0	546	16	Q97SK5 streptococc
11	176	5.0	614	2	Q8KN79 pseudomonas
12	170.5	4.9	550	2	Q8KT36 pseudomonas
13	160.5	4.6	612	16	Q8XR51 pseudomonas
14	160	4.6	1183	16	Q98QH9 mycoplasma
15	154	4.4	713	2	O05937 pseudomonas
16	145.5	4.2	1165	16	Q8NZB0 streptococc

17	140.5	4.0	1165	16	Q8K5S2 streptococc
18	139	4.0	561	16	Q9X857 streptomyce
19	139	4.0	764	16	Q92A22 listeria in
20	137.5	3.9	604	2	Q9RMC9 acinetobact
21	137.5	3.9	1165	16	Q99XX8 streptococc
22	133.5	3.8	1165	2	Q8KLP1 streptococc
23	132.5	3.8	1555	5	Q9UON0 plasmodium
24	127	3.6	560	10	Q9CS12 streptococc
25	126.5	3.6	696	16	Q9C512 arabidopsis
26	126	3.6	1179	2	Q9ACP27 streptococc
27	125.5	3.6	672	16	Q8ZBE2 streptococc
28	125	3.6	656	16	Q8CLP5 streptococc
29	125	3.6	1023	2	Q93T53 streptococc
30	124.5	3.6	1210	5	Q8IAV1 plasmodium
31	124.5	3.6	1883	16	Q9PPT2 ureaplasma
32	121.5	3.5	1088	16	Q8EUS7 mycoplasma
33	121	3.5	965	2	Q9AHL6 pasteurella
34	121	3.5	965	16	Q9CMP0 pasteurella
35	120.5	3.4	625	16	Q98R34 mycoplasma
36	120.5	3.4	992	16	O65931 mycobacteri
37	120	3.4	1046	2	O84941 streptococc
38	120	3.4	1272	16	Q98PR8 mycoplasma
39	119	3.4	776	16	Q98PT2 mycoplasma
40	118.5	3.4	579	3	Q9P6J3 schizosacch
41	118.5	3.4	839	16	Q8D518 vibrio vuln
42	118	3.4	772	2	Q46080 pedobacter
43	117.5	3.4	675	5	Q81KA2 plasmodium
44	117.5	3.4	693	5	Q94728 riptortus c
45	117.5	3.4	2902	16	Q9ZME6 helicobacte

ALIGNMENTS

RESULT 1

Q59289 ID Q59289 PRELIMINARY; PRT; 659 AA.
AC Q59289;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Heparinase III protein precursor (EC 4.2.2.8).
GN HEP.
OS Pedobacter heparinus (Flavobacterium heparinum).
OC Bacteria; Bacteroidetes; Sphingobacteria; Sphingobacteriales;
OC Sphingobacteriaceae; Pedobacter.
OX NCBI_TaxID=984;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96316388; PubMed=8702264;
RA Su H., Blain F., Musil R.A., Zimmermann J.J., Gu K., Bennett D.C.;
RT "Isolation and expression in Escherichia coli of hepB and hepC, genes
RT coding for the glycosaminoglycan-degrading enzymes heparinase II and
RT heparinase III, respectively, from Flavobacterium heparinum.";
RL Appl. Environ. Microbiol. 62:2723-2734(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Tkalec A.L.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U27586; AAB18278.1; --
KW Lyase; Signal.
FT SIGNAL 1 24 POTENTIAL.
SQ SEQUENCE 659 AA; 75806 MW; B73EDF10A1256FE2 CRC64;

Query Match 100.0%; Score 3494; DB 2; Length 659;
Best Local Similarity 100.0%; Pred. No. 9,1e-232;
Matches 659; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MTTKFKRIIVFAVIALSSGNILAQSSITRKDFDHINLEYSGLKVKVKAAGNYDDAA 60

Db 1 MTTKFKRIIVFAVIALSSGNILAQSSITRKDFDHINLEYSGLKVKVKAAGNYDDAA 60

OY 61 KALLAYREKSKAREPDFSNAEKPADIRQPIDKVTREMAADKALVHQFQPHKGYGIFDYK 120

```
Db 61 KALLAYREKSKAREPDSNAEKPADIRQPDVKVTREMDKALVHQFQPHKGYGDFYDK 120
QY 121 DINQWMPVKDNEVWQHLRVKQWQAMALVHATGDEKYAREWVYQYSDWARKNPLGLSQ 180
Db 121 DINQWMPVKDNEVWQHLRVKQWQAMALVHATGDEKYAREWVYQYSDWARKNPLGLSQ 180
QY 181 DNDKEVWMPLEVSQVSLPPTFSLFVNSPAPTAPFLMEFLNSYHQQADYLSLTHVAEQN 240
Db 181 DNDKEVWMPLEVSQVSLPPTFSLFVNSPAPTAPFLMEFLNSYHQQADYLSLTHVAEQN 240
QY 241 HRLFEAQRNLFAVGFPEFKDSPRWQRTGISVNLTEIKKQVYADGMQFELSPIYHVAID 300
Db 241 HRLFEAQRNLFAVGFPEFKDSPRWQRTGISVNLTEIKKQVYADGMQFELSPIYHVAID 300
QY 301 IFLKAYGAKRVNLKEFPQSVQTVENMINMALISISLPDYNTPMGDSWITDKNFRMAQ 360
Db 301 IFLKAYGAKRVNLKEFPQSVQTVENMINMALISISLPDYNTPMGDSWITDKNFRMAQ 360
QY 361 PASWARVPPANAIKYFATDGKQKAPNPLSKALNAGPYTFRSGWKNATVMVLKAGPP 420
Db 361 PASWARVPPANAIKYFATDGKQKAPNPLSKALNAGPYTFRSGWKNATVMVLKAGPP 420
QY 421 GEFHAQPDNGTPELFPIKGNFTPDAGVFYVSGDEAIMKLRNMYRQTRIHSITLTDNQNV 480
Db 421 GEFHAQPDNGTPELFPIKGNFTPDAGVFYVSGDEAIMKLRNMYRQTRIHSITLTDNQNV 480
QY 481 ITKARQNKWETGNNLDVLTYPNPNLDHORSVLPINKKYFLVDRAIGBATGMGVHM 540
Db 481 ITKARQNKWETGNNLDVLTYPNPNLDHORSVLPINKKYFLVDRAIGBATGMGVHM 540
QY 541 QLKESNPVFDKTKRVTYTRDGNLMQSLNADRTSLNEEGKVSYYNKLKRPAPV 600
Db 541 QLKESNPVFDKTKRVTYTRDGNLMQSLNADRTSLNEEGKVSYYNKLKRPAPV 600
QY 601 FEKPKKNAQTQNFVSIVPYDQKAPETISIRENKGNDPEKGLNLTTLINGKQQLVLP 659
Db 601 FEKPKKNAQTQNFVSIVPYDQKAPETISIRENKGNDPEKGLNLTTLINGKQQLVLP 659

RESULT 2
QBE376
ID QBE376 PRELIMINARY; PRT; 634 AA.
AC QBE376;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN GBS1885.
OS Streptococcus agalactiae (serotype III).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID=216495;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MEM316 / Serotype III;
RX MEDLINE=2242508; PubMed=12354221;
RA Glaser P., Rumiok C., Buchrieser C., Chevalier F., Frangeul L.,
RA Msadek T., Zouine M., Couve E., Lalioui L., Poyart C., Trieu-Cuot P.,
RA Kunst F.;
RT "Genome sequence of Streptococcus agalactiae, a pathogen causing
RT invasive neonatal disease."
RL Mol. Microbiol. 45:1499-1513(2002).
DR EMBL; AL766854; CAD47544.1; -
DR SAG1885; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 634 AA; 74562 MW; A31C360DC0B711D CRC64;

Query Match 7.2%; Score 250.5; DB 16; Length 634;
Best Local Similarity 22.8%; Pred. No. 8.1e-09;
Matches 127; Conservative 74; Mismatches 223; Indels 133; Gaps 21;
QY 123 NWQM-----W---PVKDN-EVRQHLRVKQWQAMALVHATGDEKYAREWVYQ 166
```

```
Db 49 NWDMEPCSKAYCLDPLEWDPVTDDPEWMLNRLQTYLPKFLVWYIVEGDKSYLRQMKYF 108
QY 167 YSDWA-----RNKPLGLSQDNDKFWRPLEVSQVSLPPTFSLFVNSPAPTAPFLMEFLN 222
Db 109 MYHWIDCOFTLKPEGA-----VSRITDTCIRCMWLKVLFLFYDFGLITEKKIKLIT 161
QY 223 SYHQADYLSLTHYABO---GNHRLFEAQRNLFAVGFPEFKDSPRWQRTGISVNLTEIKK 279
Db 162 SLREQITWRDYREKDSLSNWGILQTTAILACLYYYEDELNLPLOQFAEELLQIKL 221
QY 280 QYADGMQFELSPIYHVAIDIFLKAYGAKRVNLKEFPQSVQTVENMINMALISISLP 339
Db 222 QILDGSGVEQSIMVHVEVLKSLMELVILAPKYLPLE-----ETIEKMVTYLIAMTGP 275
QY 340 DYNTPMGDSWITDKNFRMAQFASWARVPPANAIKYFATDGKQ-----CKAPNFLSK 392
Db 276 DYCQALIGDSVDVTDTR---DILTATLVLSKSTKSFSDNVNLETLLLFQKPSIYLP 331
QY 393 ALSNA---GFYTFRSGWMD---KNATVMVLKASPPGEFHAQPDNGTPELFPIKGNFTPD 444
Db 332 EIPRATIGESAYLPFDSGHVCLDRDRIYFFKNGPFGSAHTSDNNSVCLYDKKXPIFD 391
QY 445 AGVFYVSGDEAIMKLRNMYRQTRIHSITLTDNQNVITKARQNKWETGNNLDVLTYPN 504
Db 392 AGRYTYKEEQ---LRYDFKRSSTSHSTCTLDQPLEMIK-----DSWTYN--S 433
QY 505 YPNLDH-----QRSVLFINKKYFLVIDRAIGBATGMGVHMQLKED 545
Db 434 YPKCYCQITKDRVHLVEGQLHVORASDIYVHKWLLTLPOAI----- 477
QY 546 SNPVFDKTK---NRYVTT-----YRDG---NNLMQSLNADRTSLNEEGKVSYYN 591
Db 478 -TLVIDKVCSCGEHVTNQYILDDQVYVENGFNVDKLVS-----PTTFNLEDCLSKRYN 532
QY 592 -----KELKRPAPVFE 602
Db 533 QLTESHKLKVKIKFVDE 549

RESULT 3
QBDXF4
ID QBDXF4 PRELIMINARY; PRT; 634 AA.
AC QBDXF4;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Conserved hypothetical protein.
GN SAG1897.
OS Streptococcus agalactiae (serotype V).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID=216466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2603 V/R / Serotype V;
RX MEDLINE=22222988; PubMed=12200547;
RA Tettelin H., Masignani V., Cieslewicz M.J., Eisen J.A., Peterson S.,
RA Kessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,
RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
RA Radune D., Fedorova N.B., Scanlan D., Khouri H., Mulligan S.,
RA Carty H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Maione D.,
RA Rinaudo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an
RT emerging human pathogen, serotype V Streptococcus agalactiae."
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).
DR EMBL; AE014277; AAN00759.1; -
DR TIGR; SAG1897; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 634 AA; 74622 MW; BFF2C40CDCLB84FD CRC64;
```

Query Match		7.2%;	Score 250.5;	DB 16;	Length 634;			
Best Local Similarity		22.8%;	Pred. No. 8.1e-09;	Mismatches 223;	Indels 133;	Gaps 21;		
Matches 127;		Conservative 74;	Mismatches 223;	Indels 133;	Gaps 21;			
QY	123	NWQM-----W-----PVKDN-EVRWQLHRVQWQAMALVYHATGDEKYAREWVYQ	166					
DB	49	NWDMPECSKAYCLDPLEWDKPVTDDEWLYMLNRQTYLFKFLVYVIVEGDKSYLRQMYKF	108					
QY	167	YSDWA-----RKNPLGLSQDNKDFVWPLEVDSRVQSILPTTFSLFVNSPAPTAPFLMEFLN	222					
DB	109	MYHWIDCQFTLKPEGA-----VSTIDTIGRCMSWLKVLFLDLYFGLITETKKIKLIT	161					
QY	223	SYHQADYLSHYAEQ-----GNHRLFAQRNLFAGVSFPEFKDSPRWQTGISVNLNTEIKK	279					
DB	162	SLEQITTYRDYREKDSLSNWGILQTTAILACLYYEDELNPEIQSFAEBELLQIKL	221					
QY	280	QVYADGMQFELSPIYHVAIDIFLKAYGSAKRVNLEKEFPQSYVQTVENMIMALISISLP	339					
DB	222	QILDDGSQVEQSYMYHVEVLKSLMELVILAPKYLPLE-----ETIEKMWTVLIAMTGP	275					
QY	340	DYNTMFGDSWITDKNFRMAQFASWARVPANQAIKYFATDGKQ-----GKAPNPLSK	392					
DB	276	DYCQLAIGSDVDTDR-----DILTTLATVLKSSKTSFSFSDNVNLETLLFGKPSIYLFEE	331					
QY	393	ALSNA-----GFYTFRSGWD-----KNATVMVLKASPPGEPHAQPDNGTFELFIKGRNFTPD	444					
DB	332	EIPRATIGESAYLFPDSGHVCLDRDRYIFFKNGPFGSAHSDNSVCLYDKKKFIPID	391					
QY	445	AGVFVYSGDEAIMKLRNWRQTRIHSITLTDNQMVTIKARQNKWETGNNLDVLTYNPS	504					
DB	392	AGRYTYKEEQ-----LRYDFKRSTSHSTCTLLDQPLEMIK-----DSWTYN--S	433					
QY	505	YNLDH-----QRSVLFINKKYFLVIDRAIGATGNLGVHWQKED	545					
DB	434	YPKCDYCYLTSKORYHLVGEQLHVQRASDIYHKKRWLLTLPOAI-----	477					
QY	546	SNPVDFKTK--NRVYTT-----YRDG--NNLMIQSLNADRTSLNEEGKSVYVN	591					
DB	478	TLVIDKVCSPGEHVTNQYILDQVIYENGFVNDKLSV-----PTFNLEDCLISKRYN	532					
QY	592	-----KELKRPAPVEE 602						
DB	533	QLTESHLVKKIKFVDE 549						
RESULT 4								
ID	Q8K866	PRELIMINARY;	PRT;	635 AA.				
AC	Q8K866;							
DT	01-OCT-2002 (TremBLrel. 22, Created)							
DT	01-OCT-2002 (TremBLrel. 22, Last sequence update)							
DT	01-OCT-2002 (TremBLrel. 22, Last annotation update)							
DE	Hypothetical protein SpyM3_0442.							
GN	SPYM3_0442.							
OS	Streptococcus pyogenes (serotype M3).							
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;							
OC	Streptococcus							
OX	NCBI_TaxID=198466;							
RN	[1]							
RP	SEQUENCE FROM N.A.							
RC	STRAIN=MGAS315 / Serotype M3;							
RC	MEDLINE=22133808; PubMed=12122206;							
RA	Beres S.B., Sylva G.L., Barbican K.D., Lei B., Hoff J.S.,							
RA	Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,							
RA	Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,							
RA	Schlievert P.M., Musser J.M.;							
RT	"Genome sequence of a serotype M3 strain of group A Streptococcus:							
RT	phage-encoded toxins, the high-virulence phenotype, and clone							
RT	emergence."							
RL	Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).							
DR	EMBL; AB014145; AA079049.1; -.							
KW	Hypothetical protein; Complete proteome.							
SQ SEQUENCE		635 AA;	73121 MW;	66129CA05E2E2E24 CRC64;				
Query Match		6.6%;	Score 232;	DB 16;	Length 635;			
Best Local Similarity		22.1%;	Pred. No. 1.5e-07;	Mismatches 148;	Conservative 86;	Mismatches 275;	Indels 160;	Gaps 28;
QY	64	LAYYREKSKARBEDES-----NAEKPADIRQPIDKVTREMAKALVHQFQPHKGVGYYF	116					
DB	3	LAFARFKETV-NPDFCKNLLDYQANSYAD-----OKRIADLLLTNTF-----LF	46					
QY	117	DYKGD-----INQMWPVKDNEVRWQLHRVQWQAMALVYHATGDEKY---AR	161					
DB	47	EDNWDMEPCSIPVRLTETITQKAVTDDPEWFMNLEQTYLQTLVLYLVERDERYLLTAK	106					
QY	162	EWYQVSDWA-RKNPLGLSQDNKDFVWPLEVSDR-----VQSLPTTFSLFVN-SPAFTPA	215					
DB	107	GFILNWIESAIPLDPKGLAT-----RTLDTGIRCFAWVKCL-----IYLNLFNALITQ	154					
QY	216	FLMEFLNSYHQADYLSHYAEQ-----GNHRLFAQRNLFAGVSFPEFKDSPRWQTGISV	272					
DB	155	EESLIASLEKQLQFLHNTYLDKYSLSNWGILQTTAILADAYGSDLDIAAATAPARKE	214					
QY	273	LNTEIKQVYADGMQFELSPIYHVAIDIFLKAYGSAKRVNLEKEFPQSYVQTVENMIMA	332					
DB	215	LTCQIALQILEDGSGQFEQSTMYHVEVLKSLLEL-----TALVPDYLPOLRPTLLAMSDY	268					
QY	333	LISISLPDYNFMFGDSWITDKNFRMAQFAS-----WARVPPANQAIKYFATDGKQG-	384					
DB	269	LLKMTDPDHQIPLGSDSFTDTRDILTTLAATILEEPHLKAAAPPTLD-IDSLLLLGEKGV	327					
QY	385	-----KAPNLSKALSNAFYTFRSGWDKNATVMVLKASPPGEPHAQPDNGTFELF	435					
DB	328	HIFEQLPIQTLTF-AHFEHSGHITI-----NOENYLFKNGPIGSSHTSHSDQNSICLY	382					
QY	436	IKGRNFTPDAGVFVYSGDEAIMKLRNWRQTRIHSITLTDNQMVTIKARQNKWETGNNL	495					
DB	383	YKGOPLFCADAGRYTYKEE---PLRVALKSAHSHSTAFLEEQ---LPEQIDSSWA-----	430					
QY	496	DVLTYNPSY-----PNLDHORSVLFINKKYFLVIDRAIGEA	532					
DB	431	-YLSVPKSNYCHLRQNGHVYFDGSGYQTOFSDRNNYQHDRQILILPPGIFLIIDTI--QA	487					
QY	533	TGN-----LGVHWQAKEDSNPFVDKTKNRYVTTYTRDGNLMIOSLNADRTSLNEEG	584					
DB	488	QGNHCLVSOFLDNHLEVKTD-----HLSDLRLISDY-----PFTIET	526					
QY	585	KVSYVYNKELKRPAPVFEKPKNAGTQNFVSIYPYDQGAPEISIRENKGNDFEKGKLN	644					
DB	527	ILSKYNYLTSHKLIKRPFKDKGCTS--TLVPPDDTKVTPLTPLQTKRNPETA---581						
QY	645	LTLTINGKQ 653						
DB	582	LSMHLKGRQ 590						
RESULT 5								
ID	Q9A0T7	PRELIMINARY;	PRT;	635 AA.				
AC	Q9A0T7;							
DT	01-JUN-2001 (TremBLrel. 17, Created)							
DT	01-JUN-2001 (TremBLrel. 17, Last sequence update)							
DT	01-DEC-2001 (TremBLrel. 19, Last annotation update)							
DE	Hypothetical protein Spy0628.							
GN	SPY0628.							
OS	Streptococcus pyogenes.							
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;							
OC	Streptococcus							
OX	NCBI_TaxID=1314;							
RN	[1]							
RP	SEQUENCE FROM N.A.							
RC	STRAIN=SF370 / ATCC 700294 / Serotype M1;							
RX	MEDLINE=21192684; PubMed=11296296;							
RA	Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,							


```
QY 239 GNRHLEAQRNLFAGYSEPEFKDPRWQRTGISVLNTEIKKQVYADQMPELSPIYHV-- 296
Db 181 SNWGIILQTPMLAIYHFFSKMDLEERAYHPASELKKQIETQILGDSQFEQSIYHVEV 240
QY 297 --AAIDIFLKAYGSAKRVNLEKEFFQSYVQTVENMIMALISISLDPYNTPMFGDSWITDK 354
Db 241 YKALLDLCL-----LLPDLQDSYQELLEKMKATYIQMTGLDGRTLAFGDSSTET 290
QY 355 NFRMAQFA-----SWARVFPANQAIKYPATGKQKQKAPNFKLSKALSNA 397
Db 291 TEILSLSAVLVNQEDLLNGLDVKKVLLSLFLGREKVKRLQEFKQAWQPK--SMIFEDS 348
QY 398 GFYTFRSQWGNKATVMVLKASPGEPHAQPDNGCTFEFLFKGRNFTPDAGVFVYSGDEAIM 457
Db 349 GHVCIKD----BHYLFFKNGPLGSAHSHSDENSFCLOYGQPIFIDAGRISY----- 397
QY 458 KLRNWRQTRI-----HSTLTLDNQNMVITKARQ---NKWE-----TCGNLDVLT 499
Db 398 --REIYERYLLKSAWSHSTCIVDG-----KAPERITGSWEYEPYPSLFCCHKEREQGVH 449
QY 500 YNPSY-----PNLD--HORSVLPIKXYFLVID--RAIGBATGNLGVHQLKEDSNPVED 551
Db 450 YIEGAYWSAEPDLPYLHRRKILMLVEDVWLLVDVDIRCQGO-----HEVLTQF---ILD 499
QY 552 KTKNRVYTYRDG--NNLMIQS 571
Db 500 KD-----VTYQDKINQLRLWS 516

RESULT 11
Q8KN79
ID Q8KN79 PRELIMINARY; PRT; 614 AA.
AC Q8KN79
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE Similar to prenyltransferase and squalene oxidase.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22053227; PubMed=12057956;
RA Raymond C.K., Sims E.H., Kas A., Spencer D.H., Kutyavin T.V.,
RA Ivey R.G., Zhou Y., Kaul R., Clendenning J.B., Olson M.V.;
RT "Genetic Variation at the O-Antigen Biosynthetic Locus in Pseudomonas
RT aeruginosa.";
RL J. Bacteriol. 184:3614-3622(2002).
DR EMBL; AF498414; AAM27770.1; -.
DR EMBL; AF498414; AAM27770.1; -.
KW Transferase.
SQ SEQUENCE 614 AA; 69144 MW; 2E44A45D87511F00 CRC64;
```

```
Query Match 5.0%; Score 176; DB 2; Length 614;
Best Local Similarity 23.0%; Pred. No. 0.001;
Matches 100; Conservative 55; Mismatches 178; Indels 102; Gaps 24;
```

```
QY 113 YGYEY-----GKDLNQWQWPKNEVR-----WQLHRVQWQA 146
Db 72 FGYPVYADSVPDHSNLTGVTMNSDRNW--WEIPDFDLAVGDIKAYWEASRFDW--V 127
QY 147 MALVYHATGDEKYARE---WVYQYSDWARKNPLGLSQDNDFVWR--PLEVSDRVQSLPP 201
Db 128 PALAKALAGESCALDKINAWL---NDWCNPNPPYKGN-----WKCGQEASIRVMHLAL 179
QY 202 TFSLF--VNSPAPTFAPLMEFLNSYHQADYLSHY--AEQGNHRLFEAQRNLFAGVSFP 257
Db 180 AAILLQVQVAP---PAALLDLAR--IHLRIVPTIRVAIAQDNHNGTSEAAA--LFIGGSWL 234
QY 258 EFKDSP---RWQRTGISVLNTEIKKQVYADQMPELSPIYHVAIDIFLKAYGSAKRVNL 314
Db 235 RAHGEPDAVTWMLRGRKLENRAKHLIGEDGFSQVSLNHYRVMVLDTFSLAEIWRRLQLSL 294
```

```
QY 315 EKEFP--QSYVQTVENMIMALISISLDPYNTPMFGDS-----WITDKNPR-- 357
Db 295 -VEFSARWYSRASAAATHWLHAFID-----PATGDGPNLGGANDGARLLPLTQTDYRDY 345
QY 358 -----MAQFASWARVFPANQ---AIKYFATDQKQKAPNFKLSKALSNAAGYTFRESGW 406
Db 346 RPAVOLAMALFTD--TWAYFGNCECSPRLMLGLDLFSAESPVRSRVDFDEGGFALLRSG-- 403
QY 407 DKNATVMVLKASPGEPFH--AQPDNGTFEFLFKGRNFTPDAGVFVYSGDEAIMKLRNWRQ 465
Db 404 ----QAMAMLRVPRFRPRFSQADALHLDLWLEGNLLRDAGTYSYNTESAWL---SVFPG 456
QY 466 TRHSTLTLDNQNMV 480
Db 457 TASHNTIQFDGRDQM 471

RESULT 12
Q8KI36
ID Q8KI36 PRELIMINARY; PRT; 550 AA.
AC Q8KI36
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE Similar to prenyltransferase and squalene oxidase.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22053227; PubMed=12057956;
RA Raymond C.K., Sims E.H., Kas A., Spencer D.H., Kutyavin T.V.,
RA Ivey R.G., Zhou Y., Kaul R., Clendenning J.B., Olson M.V.;
RT "Genetic Variation at the O-Antigen Biosynthetic Locus in Pseudomonas
RT aeruginosa.";
RL J. Bacteriol. 184:3614-3622(2002).
DR EMBL; AF498401; AAM27563.1; -.
DR EMBL; AF498411; AAM27714.1; -.
KW Transferase.
SQ SEQUENCE 550 AA; 61874 MW; 355BAE0565FDD86D CRC64;
```

```
Query Match 4.9%; Score 170.5; DB 2; Length 550;
Best Local Similarity 24.0%; Pred. No. 0.0021;
Matches 88; Conservative 46; Mismatches 151; Indels 81; Gaps 18;
```

```
QY 144 WQAMALVYHATGDEKYAREWVY--QYSDWARKNPLGLSQDN--DKFV----- 186
Db 80 WQA-----DRSSKLWLYNLVLDLNLNARDIG--SOPGLADKLIQSWIQANPPVSGE 128
QY 187 -WRPLEVSDRVOSLPPTFSLFVNSPAPTFAPLMEFLNSYHQADYI--STHYAEQGNHRL 243
Db 129 GWEPYPLSLRIVNLVKWLARHDERSTF-----LADSLAVQADALVQOVVEYHILGNHLP 181
QY 244 FEAQRNLFAGVSPFPFKDQSPRWQRTGISVLNTEIKKQVYADQMPELSPIYHVAIDIFL 303
Db 182 ANGKALIVFAG--AYLSGAMADRWLAKGLRLDLBELPEQLNDGGHPELSPMYHATLL----- 236
QY 304 KAYGSAKRVNLE-----KEFFQSYVQTVENMIMALISISLDPYNTPMFGDSWITDK 354
Db 237 --WDMCDLVNLTSTRSLPLDAERLPQ--WREVVVQGLKWLRSQHDPDGRISFFND----- 287
QY 355 NFRMAQFASWARVFPANQAIKYFA--TDGKQKAPNFKLSKALSNA--GFYTFRSQWGNKAT 411
Db 288 ----AAFG-----IAPEYEDIAAYAKRLDISPAHENHLLAAIYNSATGVAVALPDGVKAI 339
QY 412 VMVLKASP---PGEFHAQPDNGCTFEFLFKGRNFTPDAGVFVYSGDEAIMKLRNWRQTRI 468
Db 340 LDLAKVGDYQPG--HAHADTLSPFELSVFGRRLVYNSGTSQYGGDSE----RQRQGTAA 393
QY 469 HSTLTLT 474
Db 394 ENTUGL 399
```

Db	538	SAVLRLWELCPGAQLSGNVVQSEAGNLVTADVPRIIRIELTSGRSRYYLQTEPIPVLEV-	596
Qy	591	NKELKRP 597	
Db	597	--ELDRP 601	
RESULT 14			
Q98QH9		PRELIMINARY; PRT; 1183 AA.	
ID	Q98QH9		
AC	Q98QH9		
DT	01-OCT-2001 (TREMblrel. 18, Created)		
DT	01-OCT-2001 (TREMblrel. 18, Last sequence update)		
DT	01-OCT-2001 (TREMblrel. 18, Last annotation update)		
DE	Hypothetical protein MYPU_3820.		
GN	MYPU 3820.		
OS	Mycoplasma pulmonis.		
OC	Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.		
OX	NCBI_TaxID=2107;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=UAB CTIP;		
RX	MEDLINE=21267165; PubMed=11353084;		
RL	Nucleic Acids Res. 29:2145-2153 (2001).		
DR	EMBL; AL445564; CAC13555.1; -.		
DR	Mypulist; MYPU 3820; -.		
KW	Hypothetical protein; Complete proteome.		
SQ	SEQUENCE 1183 AA; 140128 MW; 565C3A6948E1A9D0 CRC64;		
Query Match 4.6%; Score 160; DB 16; Length 1183;			
Best Local Similarity 18.7%; Pred. No. 0.033;			
Matches 159; Conservative 131; Mismatches 252; Indels 308; Gaps 44			
Qy	7	KRII--VEAVIAL-----SSGNTLAQSSSITR-----KDFD-----	35
Db	4	KIIASSLFLPMSLFLFLSSNTSTIAQENITQVFKNSLENFDEYRKNNDDEEIGEKTLS 63	
Qy	36	-----HINLSEGLEKYNKVAAGNYDDAAKALLAYYREKSAREPDFSNAEKPAD 86	
Db	64	IVSKFIETKNLNIYKSDFEKINYIVNDSNE-----NFSLGNNSQ 105	
Qy	87	IRQPIDKVTEMADKALVHQPQPHKGYGFDYDGKDINQMWPVKDNEVRWLHRVKKWQA 146	
Db	106	IE-----NMTNLNVKTNIDID-KLNSVKFF-- 128	
Qy	147	MALVYHATGDEKAREWYQYSDWARKNPLGLSQNDKFVWRPLEVSRVQSLPFTFSLF 206	
Db	129	-----FNDKSQSW-----DFIEITNIQRIEKIKDLSQK 160	
Qy	207	VNSPAFTA-----FLME-----FLNSYHQADYL-----STHYAQGNHRLFPAQ 247	
Db	161	YNDNHFSVSIKKLDIFYDYKSPKNSLFLNEFRKKVDVILBKRVNVFYKKEINSIQIV 220	
Qy	248	RNLFAGVSFFPEFKDSPR--NRQ-TGISVLNTEIKKQVYA--DGMQFELSLPIYHVAADI 301	
Db	221	KD-FDKLNFELTSSSLFLFKWKNVNDLLHNEIYKIQALDFNKKKFEIK--YQLISDDL 277	
Qy	302	FLKAYGSAKRVNLEKFPQSY--VOTVENMTMALISLDPYNT-----	343
Db	278	I-----SSDVKFLLLFRGEYFNQNIIDLGIKVKISRSINSPIKQRLLIYYKKNV 331	
Qy	344	-----PMF-----GDSWITDKNFRMAQFASWAR---VFPAQAQIK 375	
Db	332	YKNGDKNGKEKVPQFIHEEREISNEYGGHVLVNOQFFRMAFISDQNEVIYVDIAD 391	
Qy	376	-----YFATDGKQKAPN-----FLSKALSNAAGFTFRSGWD--KNA 410	

RESULT 13	Q8XR51	PRELIMINARY;	PRT;	612 AA.
ID	Q8XR51	PRELIMINARY;	PRT;	612 AA.
AC	Q8XR51;			
DT	01-MAR-2002	(TREMUREl. 20, Created)		
DT	01-MAR-2002	(TREMUREl. 20, Last sequence update)		
DT	01-MAR-2002	(TREMUREl. 20, Last annotation update)		
DE	Hypothetical protein Rsp1010.			
GN	RSP1010 OR RS02344.			
OS	Ralstonia solanacearum (Pseudomonas solanacearum).			
OC	Plasmid megaplasmid.			
OC	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;			
OC	Ralstoniaceae; Ralstonia.			
OX	NCBI_TaxID=305;			
RN	[1]			
RC	SEQUENCE FROM N.A.			
RP	STRAIN=GMI1000;			
RX	MEDLINE=21681879; PubMed=11823852;			
RA	Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,			
RA	Ariat M., Ballaut A., Brottier P., Camus J.C., Cattolico L.,			
RA	Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,			
RA	Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,			
RA	Siguer P., Thebaud P., Whalen M., Wincker P., Levy M.,			
RA	Weissenbach J., Boucher C.A.;			
RT	"genome sequence of the plant pathogen Ralstonia solanacearum.";			
RL	Nature 415:497-502(2002).			
RL	EMBL; AJ646082; CAD18161.1; -			
KW	Plasmid; Hypothetical protein; Complete proteome.			
QY	SEQUENCE 612 AA; 68273 MW; 78FCB9B4C8CD4277 CRC64;			
Query Match	4.6%; Score 160.5; DB 16; Length 612;			
Best Local Similarity	21.3%; Pred. No. 0.012; Indels 119; Gaps 31;			
Matches 129; Conservative 85; Mismatches 274;				
QY	49 KAVAAGNYDDAAKALLA-----YYREKSKAREPDF-SNAEKPADIROPIDKVTREMDAKA 102			
Db	56 KAAPEGLVGHAVEETVAVRVFGWYQPAQA-EPDHRNPFNGARVNEP-SQPMWQIAD-- 111			
QY	103 LVHQFQPHKGYPDYGDKDNWQMPVKDNEVRQLHRVKKWQAMALVYHATGDEKVA-- 160			
Db	112 ----FDPAVG-----DIIKAIWEASRFDW--VIVLIAQQAVRGRPOAMT 147			
QY	161 --REWVYQYGDWARKNPLGLSQDNDKFWR-PLFVSDRVQSLPPTFSLVFVNSPAFTPAFL 217			
Db	148 QLANWL---TSWARANPPYLGPN-----WKCGQEAALRVHMLAALVILGHFSAAPA-L 198			
QY	218 MEFNLSVHQQ-ADYLSTHYABQGNHRUFEAQRNLFAGVSPFEFKDSP---RWRGTGTSVL 273			
Db	199 MALVVRTHLQRIAPTGLVAYIADQNNHGTSEAAA-LFIGSGWLAACQDPDGRHWQAGSHWL 257			
QY	274 NTEKKQVYADGMQFELSPYVHVAIDIFLKAYGSAKVNLEKEFP-----QSVQVQIVE 327			
Db	258 ENRARKLIADGSGFSQHSVMYH-----RLMLDITYSNAEVRWRHWSLPAFSAQLQARLGGAS 313			
QY	328 NMIVALISILPDYNTPMFGDS-----WITDKN-----FRWAQFASKARV 367			
Db	314 NWLYQMIDRTTGD--APNLGANDGARLLPLTATDHRDFRPSVQLACALFORADAFGWDGE 371			
QY	368 FPAQAIAKYPATDGKQKAPNFLSKALSAGFYTFRSGWDKKNATVMVLKASPPGEFH--A 425			
Db	372 W--SDAURLVGLVPRPEQVRPPARTHMEAGYGILLRYG---RAPALF---NLPHRRHPS 423			
QY	426 QPDNGTPELETIKGRNFTPDGAFVYVSGDEATMLKLNRYQTRIHSTLTLDNQNMV--ITK 483			
Db	424 QADALHVDVFWLGGKNNLRDAGSFSYASAES---AGVFSGTASHTVQFDQDQMPRLSR 479			
QY	484 ARQKQWETGNLVDVITNPN-----SYPNLDHQBSVLFINKKFLVLDIDRGAIGATG 534			
Db	480 FLFGAWLTKARDVEPVKQTADGVTCAAGYRVQOGASHIRA-LTLGKRSLRVYDR-VGGFRR 537			
QY	535 NLGVHWOL----KEDSNPVFOKTKNRYVTTVTRDGNLMIQSLNADRTSLNEEGKVSVYV 590			

Db 392 NIDNLEFFDLEHQQRKKKTKITIKTKDNTKNTLYEIIYSKELKDSALIKARWENWDPINNL 451
QY 411 TVMVL-----KASPPGFHQAQDNGTGFELFI-----KGRNFTPD-----AGVF 448
Db 452 EHKLVKDNKNGKVNPKYINPRINPHNGLEEKIFWLEHEKLNHLSNFKKVFNIKEGAF 511
QY 449 V-----YSGD-----EAIMKLNRYQRIRHSTLTLDNONNVITKARONKW-ETGNL 495
Db 512 VKYSISQKAYSIDLPPENSLTLKKYFWNL-----SNDDFSOEELITIKDNNNSFSKGNFL 568
QY 496 DVITYN-----PSYNLDHQRSLVLFINKYFLVIDRAIGEAATGNLGVHWOLK--- 543
Db 569 LELKNTNYFNSYNLISVGYSNLKHNLDEL--IDKKILVPVEKSI---AGNILKNYLLKNFN 623
QY 544 -EDSNPFVDKTKRVVITYYEDGNLMQISNAORTSLNEZ-----EGKV-----SYVYNKE- 593
Db 624 FEES--ISSLSYBEIVQNVKLLNVNIEFLNKNKYNKKIFDIENKMDSLSYVDYVFNOK 681
QY 594 ----LKRPAFVFEKPKKNAGTQNFVSIVPYDQKAPETISIRENKGNDPEKGLN---LT 646
Db 682 LISLDEEFKINSFKN--TKNFVEIDYQYD---HPNYDLTFEKSIFYKDKINKKQIT 736
QY 647 LTINGKQQLV 656
Db 737 ININ-KERII 745

RESULT 15
O05937 PRELIMINARY; PRT; 713 AA.
AC O05937;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Alginatase lyase.
GN ALXLL.
OS Pseudomonas sp. (strain OS-ALG-9).
OC Bacteria; Proteobacteria.
OX NCBI_TaxID=86038;
RN [1]__
RP SEQUENCE FROM N.A.
RC STRAIN=OS-ALG-9;
RA Kraiwattapanong J., Ooi T., Kinoshita S.;
RT "Nucleotide sequence and expression of a gene (alyII) for an alginatase lyase from Pseudomonas sp. OS-ALG-9";
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB003330; BAA19848.1; -;
KW Lyase.
SQ SEQUENCE 713 AA; 79805 MW; 2CE6CD2895899F2E CRC64;

Query Match
Best Local Similarity 4.4%; Score 154; DB 2; Length 713;
Matches 153; Conservative 92; Mismatches 234; Indels 276; Gaps 42;

QY 87 IQPIDKVTREMADKALVHQF---QPHKGYGDFYDKDINWQWPPVKDNEVRWQLHRVKW 143
Db 48 LRQSYQAV-KNAADKALAQPIVVPVPEKGGGVTH-----EQHKNY 88
QY 144 WQAM-ALVYHATGEKYA-----REWYQYSDWARKNPLGLSDN-----DKF 185
Db 89 SNMLNGVAYQISGEKKYADYKVNVMNLNYSQYQKWP-LHPKRKSEEDGGRIFWQSLNDF 147
QY 186 VNR--PLEVSDRV-QSLPPTFSLFVNSPAPFTAPFLMEF-----LNSYHQADYLSTH 234
Db 148 VWQLYTIQAYDLVYDIPADTKRTIEKLFVP--ILKFTEDRYDVFNKIH----- 196
QY 235 YAEQGNHRLFEAQRNLFA-GVS-----FPE--FKDSPRWRQTGISVLNTEIKKQVYA 283
Db 197 -----NHGTW-----NLAAGITCYVLNKEEYVEMAIKSGKDKGTGL-----AQIDQLFSP 244
QY 284 DGMQFELSPYHVAIDIFPKAYGSAKRVNLEKEPPQSYVQTVENMIMALISISLPDYNT 343
Db 245 DGYNME-GPYQRYALLPFVL---FAKAIN-NVEPSRKIFEYRDKLLSKAIHTSL----- 294

QY 344 PMEGDSWITDKNFRMAQFASWARVFPANQAIK----- 375
Db 295 ----QTSYTDKTF-----FPLNDAIKDKTYVESVELVYGVDLAYADIKAEVDLLD 339
QY 376 -----YFATDGKQKAPNELSKAL-----SNAGFYTFERSGWDKNATV 412
Db 340 IARQONRVIVSDAGLKVAADLAAGKAVPPKYQTLWIRDDGKGDEGGGLGILRNGENTDQOC 399
QY 413 MVLKASPPGGEFHAQPD-----NGTFELFIKGRNFTPDAG-----VFVYSGDEAIMKL 459
Db 400 VVLKAASOGMGHGFDRLLNLLFYDNTTEIF-----PDYGAARFLNIITKNGGGYLPEN 452
QY 460 RNWYRQTRIHSTLTLD-----NQNMVITKARONKMET----- 491
Db 453 NTWAKQTVAHNALVVDQTSHFNAKLGPADKASPTLLYFNSQPNLKVVSAKEDKAYTDVTM 512
QY 492 -----GNLADVLTNTSPNLDHORSVLF-----INNKYFLVID-----PAIGE 531
Db 513 LRISALVKVEGLDKPLIIDVMQAQSAKSHQYDLPFWYKQOLVNTSPPTAKANQLTALGD 572
QY 532 ATGNLGVHWQLKEDSNPFVDKT-----KNRVYTTYTRDGN--LMIOQLNADRTSLNE 581
Db 573 KNGYQHI-W--LNASNPLEGKSGMVGLLNKNRYTTHFVSDNPLEVKVLLSIGANDEPMNL 629
QY 582 BEGKUSYVYNKELKRPAPFVEKPKKNAGTQNFVSIVPYDQ-----OKAPEIS--- 629
Db 630 VDGK-----AFMLSSGQN---QTFVSIETHTGGTDPINETVSSALPTVSGLK 674
QY 630 -IRENKGNDFEKGLN--LTLTIN--GKQOLVLV 658
Db 675 LIKSDAQOTIISFKVNERITYTYQINTEKQOQYII 709

Search completed: July 28, 2003, 20:52:10
Job time : 114 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 28, 2003, 20:31:37 ; Search time 26 Seconds
(without alignments)
1191.946 Million cell updates/sec

Title: US-09-802-285A-2

Perfect score: 3494

Sequence:

1 MTTKIFKRIITVFAVIALSSG.....KGKLNLTITNGKQLVLP 659

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	123	3.5	532	1 INV2 YEAST	P00724 saccharomyc
2	120	3.4	677	1 SACC_BACSU	P05656 bacillus su
3	119	3.4	1138	1 C7AB BACUA	Q45707 bacillus th
4	117.5	3.4	995	1 DPO1 KLULA	P09804 kluyveromyc
5	116	3.3	4563	1 APB HUMAN	P04114 homo sapien
6	115.5	3.3	1113	1 Y14D MYCPN	P75033 mycoplasma
7	115.5	3.3	1314	1 SS22 YEAST	P25390 saccharomyc
8	115	3.3	1063	1 HGPI HAEIN	P44795 haemophilus
9	114.5	3.3	1167	1 CLAA BACTU	P56956 bacillus th
10	114.5	3.3	1211	1 EX5B HAEIN	P45157 haemophilus
11	113.5	3.2	1398	1 PLS PYRFU	P72186 pyrococcus
12	113	3.2	2179	1 POLG EC23W	O73556 e genome po
13	112	3.2	532	1 INV1 YEAST	P10594 saccharomyc
14	112	3.2	532	1 INV4 YEAST	P10596 saccharomyc
15	112	3.2	1138	1 C7AB BACUK	Q45708 bacillus th
16	111.5	3.2	1002	1 TAGA VIBCH	P24019 vibrio chol
17	110.5	3.2	473	1 SABC BACST	P94468 bacillus st
18	110.5	3.2	850	1 PRSN PIERA	Q9u8q4 pteris rapa
19	109.5	3.1	965	1 AMPN RABIT	P15541 oryctolagus
20	109.5	3.1	2334	1 WAPA BACSU	Q07833 bacillus su
21	109	3.1	951	1 HEX AD805	P04133 human adeno
22	106.5	3.0	819	1 CSDI ECOLI	P93736 escherichia
23	106.5	3.0	850	1 PRSN PIERB	O59736 pteris bras
24	106.5	3.0	1104	1 COLA CLOPE	P43153 clostridium
25	106	3.0	1132	1 TERT OXYTR	O76332 oxytricha t
26	106	3.0	1455	1 GTFC STRMU	P13470 streptococ
27	105	3.0	473	1 SABC BACSU	P05655 bacillus su
28	105	3.0	529	1 UDBB HUMAN	O75310 homo sapien
29	105	3.0	1042	1 TIRH METJA	O60295 methanococ
30	105	3.0	1180	1 PC12 MOUSE	O55134 mus musculus
31	105	3.0	1271	1 Y338 MYCGE	P47580 mycoplasma
32	104.5	3.0	1207	1 DPOL ASF7B	P42489 african swi
33	104.5	3.0	1750	1 Y832 METJA	Q58242 methanococ

34	104	3.0	537	1 TEE6 STRPY	P18481 streptococ
35	104	3.0	802	1 CSD2 ECOLI	P33513 escherichia
36	103.5	3.0	543	1 APPA BACSU	P42061 bacillus su
37	103.5	3.0	1183	1 CNA STAAU	Q53654 staphylococ
38	103.5	3.0	1433	1 SUBE BACSU	P16397 bacillus su
39	103	2.9	674	1 ETE1 MSEPV	Q9yvx9 melanoplus
40	103	2.9	767	1 TOP1 MOUSE	Q04750 mus musculu
41	103	2.9	837	1 HFC1 HAEIN	P33397 haemophilus
42	103	2.9	917	1 SYI STAAU	P41972 staphylococ
43	103	2.9	1034	1 BGAL BACWE	O52847 bacillus me
44	102.5	2.9	2136	1 YCF2 MARPO	P09975 marchantia
45	102	2.9	513	1 YMM6 YEAST	Q03104 saccharomyc

ALIGNMENTS

RESULT 1
INV2 YEAST STANDARD; PRT; 532 AA.
AC P00724;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Invertase 2 precursor (EC 3.2.1.26) (Beta-fructofuranosidase 2)
DE (Saccharase).
GN SUC2 OR YIL162W.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
CX NCBI_TaxID=4932;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=83168934; PubMed=6300785;
RA Taussig R., Carlson M.;
RT "Nucleotide sequence of the yeast SUC2 gene for invertase.";
RL Nucleic Acids Res. 11:1943-1954(1983).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX PubMed=9169870;
RA Churcher C.M., Bowman S., Badcock K., Bankier A., Brown D.,
Chillingworth T., Connor R., Devlin K., Gentles S., Hamlin N.,
Harris D.E., Horsnell T., Hunt S., Jagels K., Jones M., Lye G.,
Moule S., Odell C., Pearson D., Rajandream M.A., Rice P., Rowley N.,
Skellton J., Smith V., Walsh S., Whitehead S., Barrell B.G.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome IX.";
RL Nature 387:84-87(1997).
[3]
RN SEQUENCE OF 1-30 FROM N.A.
RX MEDLINE=85137467; PubMed=6396505;
RA Sarokin L., Carlson M.;
RT "Upstream region required for regulated expression of the glucose-repressible SUC2 gene of Saccharomyces cerevisiae.";
RL Mol. Cell. Biol. 4:2750-2757(1984).
[4]
RN SEQUENCE OF 1-21 FROM N.A.
RX MEDLINE=87064535; PubMed=35377118;
RA Kaiser C.A., Botstein D.;
RT "Secretion-defective mutations in the signal sequence for Saccharomyces cerevisiae invertase.";
RL Mol. Cell. Biol. 6:2382-2391(1986).
[5]
RN SEQUENCE OF 21-532, AND CARBOHYDRATE-LINKAGE SITES.
RX MEDLINE=88213364; PubMed=3284881;
RA Reddy V.A., Johnson R.S., Biemann K., Williams R.S., Ziegler F.D.,
Trimble R.B., Maley F.;
RT "Characterization of the glycosylation sites in yeast external invertase. I. N-linked oligosaccharide content of the individual sequons.";
RL J. Biol. Chem. 263:6978-6985(1988).
[6]
RP ACTIVE SITE ASP-42.

RX MEDLINE=90293006; PubMed=2113524;
RA Reddy V.A., Maley F.;
RT "Identification of an active-site residue in yeast invertase by
RT affinity labeling and site-directed mutagenesis.";
RL J. Biol. Chem. 265:10817-10820(1990).
CC - CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-
CC fructofuranoside residues in beta-D-fructofuranosides.
CC - PFM: THE ENZYME EXISTS IN TWO STATES: A GLYCOSYLATED EXTERNAL
CC FORM, OR A NON GLYCOSYLATED INTERNAL FORM.
CC - MISCELLANEOUS: THE SEQUENCE OF THE SECRETED FORM OF THE ENZYME IS
CC SHOWN.
CC - SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC
CC EMBL; Z46921; CAAB7030.1; -
CC EMBL; V01311; CA24618.1; -
CC EMBL; K03294; AAA35127.1; -
CC EMBL; M13627; AAA35129.1; -
CC EIR; A00899; IFBY.
CC GlycosuiteDB; P00724; -
CC COMPILEYEAST-2DPAGE; P00724; -
CC SGD; S0001424; SUC2.
CC GO; GO:0005737; C:cytoplasm; IDA.
CC GO; GO:0005576; C:extracellular; IDA.
CC GO; GO:0004564; F:beta-fructofuranosidase activity; IMP.
CC InterPro; IPR001362; Glyco_hydro_32.
CC Pfam; PF00251; Glyco_hydro_32; 1.
CC SMART; SM00609; GLYCOSYL_HYDROL_F32; 1.
CC PROSITE; P500609; GLYCOSYL_HYDROL_F32; 1.
CC Hydrolase; Glycosidase; Glycoprotein; Multigene family; Signal.
FT SIGNAL 1 19
FT CHAIN 20 532 INVERTASE 2 (EXTERNAL).
FT CHAIN 21 532 INVERTASE 2 (INTERNAL).
FT ACT_SITE 42 42
FT CARBOHYD 23 23 N-LINKED (GLCNAC. . .).
FT CARBOHYD 64 64 N-LINKED (GLCNAC. . .) (PARTIAL).
FT CARBOHYD 97 97 N-LINKED (GLCNAC. . .).
FT CARBOHYD 111 111 N-LINKED (GLCNAC. . .).
FT CARBOHYD 118 118 N-LINKED (GLCNAC. . .).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (PARTIAL).
FT CARBOHYD 266 266 N-LINKED (GLCNAC. . .) (PARTIAL).
FT CARBOHYD 275 275 N-LINKED (GLCNAC. . .) (PARTIAL).
FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .).
FT CARBOHYD 369 369 N-LINKED (GLCNAC. . .).
FT CARBOHYD 384 384 N-LINKED (GLCNAC. . .).
FT CARBOHYD 398 398 N-LINKED (GLCNAC. . .).
FT CARBOHYD 512 512 N-LINKED (GLCNAC. . .) (PARTIAL).
FT MUTAGEN 42 42 D - N: LOSS OF ACTIVITY.
FT CONFLICT 409 409 A -> P (IN REF. 5).
SQ SEQUENCE 532 AA; 60639 MW; 7D8AB33B672B775 CRC64;

Query Match 3.5%; Score 123; DB 1; Length 532;
Best Local Similarity 18.9%; Pred. No. 0.46;
Matches 109; Conservative 76; Mismatches 179; Indels 214; Gaps 31;

QY 94 VTRMADKALVHOFQPHKGY-----GYFDYGDKNWOMW-PVKDNEVRWQLHRVKNWQAM 147
DB 21 MTNETSDRLVH-FTPNKGWMDNDGLWYDEKAKHLYFOYNPNTVW--GTPLEFW-- 75
QY 148 ALVYHATGDEKYAREWYQYSDWARKNPLGLSQDNQDKFVWRPLEVSDRVQSLPTPTSLFV 207
DB 76 ----HATSDD-----LTNW-EDQPIALPKRN----- 97
QY 208 NSPAFTPAFLMEF-----LSNYHQADYLSTHYAEQGNHLEF 245
DB 98 DSGAFSGSMVVDYNNYSGFFNDTIDPRQRCVAVIWTYNTPESEQYIS--YSLDGGYTFTE 155

QY 246 AQRNLFAGVSPPEFKD-----SPRWRTGISVLNTEIKQVYAD----- 284
DB 156 YQKPVLAANSTQFEDPKVFWYEPQKWMIT--AAKSQDYKIEIYSSDDLKSWKLESAPA 213
QY 285 -----GMQFEL-----SPIYHVAIDI-----FLKAY 306
DB 214 NEGFUGYQECPLTEVTEQDPSPKSYWMTFISNPGAPAGGSFNOYFVGSFNGTHFEAF 273
QY 307 GSAKVNLEKEFPQSY--VOTVENMIMALISLSLDYNTPMFGDSWITDKNPRMAQFA-- 362
DB 274 DNQSRV---VDFGKDYALQTFN-----TDTYGSAL-GIAWAS--NWEYSAFVPT 319
QY 363 -----SWARVFPANQAIKYFA---TDGKQKAPNFLSKALSNAAGFTFRSGWDKNATV 412
DB 320 NPWRSSMSIARVRFSLN--TEYQANPETELINLKABPILN--ISNAG-----PWSRFATN 369
QY 413 MYLKASPGEGEHAQPDNGTFFELFKGRNFTPDAGVVFVSGDEAIMK-----LRNWRQTR 467
DB 370 TTLTKANSYVNDLSNSTGTLEPEL-----VYAVNTTQTISKSVFADLSLWFKGLE 419
QY 468 -----IHSTLTLDNQNMVITKARONKWETG---NNLDVLTVTNPSYPN----- 507
DB 420 DPEEYLRMGFEVSASSFFLDKRGNSKVKFKENPYFTNMSVNNQPFKSENDLSYKVKYGL 479
QY 508 LDHQRSVLFIN-----KKYFLVIDRAIGETAGNLGV 538
DB 480 LDQNIILELYFNDGDVVNTYEMTTGNALGSVNMITGV 517

RESULT 2
SACC BACSU STANDARD; PRT; 677 AA.
ID SACC BACSU STANDARD; PRT; 677 AA.
AC P05656;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Levanase precursor (EC 3.2.1.65) (2,6-beta-D-fructan
DE fructanohydrolase).
GN SACC.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87286401; PubMed=3112519;
RA Martin I., Debarbouille M., Ferrari E., Klier A., Rapoport G.;
RT "Characterization of the levanase gene of Bacillus subtilis which
RT shows homology to yeast invertase.";
RL Mol. Gen. Genet. 208:177-184(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=88067786; PubMed=3120151;
RA Schoegendorfer K., Schwab H., Lafferty R.M.;
RT "Nucleotide sequence of a cloned 2.5 kb PstI-EcoRI Bacillus subtilis
RT DNA fragment coding for levanase.";
RL Nucleic Acids Res. 15:9606-9606(1987).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=97286548; PubMed=9141695;
RA Parro V., San Roman M., Galindo I., Purnelle B., Bolotin A.,
RA Sorokin A., Mellado R.P.;
RT "A 23911 bp region of the Bacillus subtilis genome comprising genes
RT located upstream and downstream of the lev operon.";
RL Microbiology 143:1321-1326(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,

Query Match 3.4%; Score 120; DB 1; Length 677;
 Best Local Similarity 18.8%; Pred. No. 1;
 Matches 123; Conservative 83; Mismatches 216; Indels 232; Gaps 32;

QY	41	YSGLEKYNKAVAGNYDDAAKALIAYY---REKSKAREPDFSNAEKPADIRQPIDKVTRE	97
DB	105	PSGSAVDKNNKTSFGQTGKEKPLVAIYTDQREGHQVQSIAYSNDK-----GRWTWK	155
QY	98	MADKALVHQPHQKHGYCYFDYGDINDNMWPFVKDNEVRWLHVKWQWAMA-----LVYH	152
DB	156	YAGNPVIY-----PNEG-----KXDF-----RDPKVFWYEBKKWVMVLAAGDRILIY-	197
QY	153	ATGDEKYAREWVTOYSDWARKNPLGLSQDNDFKWRPRLPVSQVSLPPTFSFLFVNSPAF	212
DB	198	---TSKNLKKQWY-----ASEFGDQSGHGVW-----ECPDLFELPVDGNPN	237
QY	213	TPAFLMEF-----LNSYHQADYIYJTHYAEQG--NHRLLF-EAQRNLFAGVSFFPEF	259
DB	238	QKKWVMQVSGNGAVSGSGMQYFVGDFDGTHTFKNENPNKVLWTDYGRDYFAAASWSDI	297
QY	260	--KDSPR-----WRGTGISVLNTEIKKQVYADGMQFELSPIYHVA	298
DB	298	PSDTSRLKWLGMNSWQYANDVTPSRSA--TSIPRELKLKAFTEGVRVVQTPVKELET	355
QY	299	I-----DIFLKAYSAGKVNLE-KEFPQS-----YVQTVENMTM---	331
DB	356	IRGTSKKKWNLITSPASHNNLAGSGDAYEINAEFKVSPGSAAEFGFKVRTGENQFTKVG	415
QY	332	-----ALISLSLDYNTPMFGDSWITDKN---FRMAQFASWAVFPANQAIKFTATDGKQ	383
DB	416	YDRNNAKLFVDRSESGNDTFNPAFTGKETAPLKPVGNGVKLRIFFDRSSVEVFGNDGKQ	475
QY	384	GKA---PNFLSKAL-----SNAG-----FYTFRSGWDMKNATVMVLK	416
DB	476	VITDIILPDRSSKGLIYAANGVKVKSITIHLKKWGTTFPMSNMTGW-----TTV-----	528
QY	417	ASPPGFHQAQPDNGTFELFIKGRNFTPDAGVFYIS-----GDEA	455
DB	529	-----NGTWADTIEGQGRSGDGSFILSSASGSDFTYESDITIKDNGRFGAGA	576
QY	456	IMKLRNWRQTRIHSTLITLDNQNWITKARQNKWETGNLNDLVITYNPSYPLNDHORSVL	515
DB	577	LM-----FRSDKDAKNGYLANVDKHDLVKFFKFPENGAAASVIAEYKTP-----I	620
QY	516	FINKKYFL-----VIDRAIGE-----TGNLGVH-WOLKEDSNPVF	550
DB	621	DVNNKYHLKTEABGDRFKIYLDLRLVIDAHDVSFSEGGQFGLNVW-----DATAVF	670

RESULT 3
 C7AB_BACUA STANDARD; PRT; 1138 AA.
 ID_C7AB_BACUA
 AC Q45707;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Pesticidal crystal protein cry7Ab (Insecticidal delta-endotoxin
 DE CryVIIA(b)) (Crystalline entomocidal protoxin) (130 kDa crystal
 DE protein)
 GN CRY7AB OR CRYVIIA(B).
 OS Bacillus thuringiensis (subsp. dakota).
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 NCBI_TaxID=132268;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HD511;
 RA Payne J.M., Fu J.M.;
 RT "Coleopteran-active Bacillus thuringiensis isolates and genes encoding
 RT coleopteran-active toxins."
 RL Patent number US5286486, 15-FEB-1994.
 CC -!- FUNCTION: PROMOTES COLLOIDSMOTIC LYSIS BY BINDING TO THE MIDGUT
 CC EPIITHELIAL CELLS OF COLEOPTERA.
 CC -!- DEVELOPMENTAL STAGE: THE CRYSTAL PROTEIN IS PRODUCED DURING

CC SPOULATION AND IS ACCUMULATED BOTH AS AN INCLUSION AND AS PART OF
CC THE SPORE COAT.
CC -!- MISCELLANEOUS: TOXIC SEGMENT OF THE PROTEIN IS LOCATED IN THE N-
CC TERMINUS.
CC -!- SIMILARITY: BELONGS TO THE DELTA ENDOTOXIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U04367; AAA21120.1; --
DR HSP; F07130; IDJC.
DR InterPro; IPR001178; Endotoxin.
DR InterPro; IPR005638; endotoxin_C.
DR InterPro; IPR005639; endotoxin_N.
DR Pfam; PF00555; endotoxin; 1.
DR Pfam; PF03944; endotoxin_C; 1.
DR Pfam; PF03945; endotoxin_N; 1.
KW Toxin; Sporulation.
KW SEQUENCE 1138 AA; 129778 MW; 01DF7072C074CE88 CRC64;
Query Match 3.4%; Score 119; DB 1; Length 1138;
Best Local Similarity 19.8%; Pred. No. 2,5;
Matches 132; Conservative 85; Mismatches 206; Indels 244; Gaps 39;
QY 48 NKAVA-----AGNYDAKALLAYREKSKAREPDPFSAEKPADIPQIDKVTREMAKAL 103
DB 121 NKAALEGLGN-----LTIYQ-----ALEDLNPPDPATITRVIDRF--RIID-AL 167
QY 104 VHQFQPHKGYGFD-----YKQDINQM-----WPKQNEV----- 134
DB 168 FESYMFSPFVAGYEIPLLTVYAQAANLHALLRDLSTLYGDRKMGFTONNTEENYNQKKHI 227
QY 135 -RWQLHRVQWQA-MALVVHATGDE-----KVAREWVYQSDWAR-----KNPLGLSDND 183
DB 228 SEYSNHCWVYNSGLRINGSITVEQWYNNRFRERIMLVLDIAAFVPIYDFRMYSWEITS 287
QY 184 KFWR-----PLEVSRVOSLPPTFLFVNSPAFTPAFLMEFLNSVHQADYLSHYASQ 238
DB 288 TQLTREVYTDPTSLISINPDIGFSQMENTAPRTPHLV-----DYLDLYIYT 336
QY 239 GNRHLF--EAQRNLF---AGVSFPEFKDSPRWRQTGISVLNTEIKKQVYADGMQFELSP 292
DB 337 SKYAFSHBIQDPLFYWCVHKVSFKKSEQSNLY--TTGI-----YKTSYGVISG 384
QY 293 IYHVAIDIF-----LKAYGSAKRVNLEKEFPQSYVQTVENMIM 331
DB 385 AYSFRGNDIYRTLAAPSVVYPYQYGVQVEFYGVKGHVHYRGD--NKYDLTYDSIDQ 442
QY 332 -----ALISLDPVNT-----PMFGDSWITDKNFRMAQFASWARVFP 369
DB 443 LPDPGEPIHEKYTHRLCHATAISKSTPDYDNATIPF--SW-----THRSAEY--YNNRIY 494
QY 370 AN-----QATKYPATDCKQG--KAPNLSKALSNAAGYTFPSGW--DKNATVMVLKASPPG 421
DB 495 NKIKKIPAVKMKLDDLSTVKGFGFTGGDLVVRG-----SNGYIGIKATV-----NSPLS 546
QY 422 E-----PHAQPDNGTTFELFKGRNFTPDAGVFVYSGDEAIMKLRNWRQTRIHSITLTD 475
DB 547 QKYRVVRVYATSVGLFNFI-----TNPSYPN-----LDH--QRSVL 515
QY 476 QNMVITKARQNKWET--GRNLDVITY-----LDH--QRSVL 515
DB 568 NDEIALQKNFQSTVETIGCKD-LTVGSGFYIEYSTIOPNEBPKPITLHNLHNSNPF 626
QY 516 FINKKPLVLDRAIGATGNLGVHMQKEDSNPVDKTRVYTYRDGNLNM---IQSL 572
DB 627 YVDSIEFIPVD-----VNYDEKEK-----LEKAKAVNLTFTEGNALQKVYTDY 671

QY 573 NADRTSL 579
DB 672 KVDQVSI 678
RESULT 4
ID DPO1_KLU1A STANDARD; PRT; 995 AA.
AC P09804;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE DNA polymerase (EC 2.7.7.7).
GN RFI.
OS Kluyveromyces lactis (Yeast).
OG Plasmid pGKL-1.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID=28985;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS 2359/152;
RA Sor F., Fukuhara H.;
RT "Structure of a linear of the yeast Kluyveromyces lactis;
RT compact organization of the killer genome.";
RL Curr. Genet. 9:147-155(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=84297209; PubMed=64730399;
RA Stark M.J.R., Mileham A.J., Romanos M.A., Boyd A.;
RT "Nucleotide sequence and transcription analysis of a linear DNA
RT plasmid associated with the killer character of the yeast
RT Kluyveromyces lactis.";
RL Nucleic Acids Res. 12:6011-6030(1984).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=88124274; PubMed=3340559;
RA Stark M.J.R.;
RT "Resolution of sequence discrepancies in the ORF1 region of the
RT Kluyveromyces lactis plasmid kl.";
RL Nucleic Acids Res. 16:771-771(1988).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=85037931; PubMed=6387625;
RA Tokunaga M., Kawamura A., Hishinuma F.;
RT "Cloning and nucleotide sequences of the linear DNA killer plasmids
RT from yeast.";
RL Nucleic Acids Res. 12:7581-7597(1984).
RN [5]
RP POSSIBLE IDENTIFICATION OF PROTEIN.
RX MEDLINE=88067726; PubMed=3684586;
RA Jung G., Leavitt M.C., Ito J.;
RT "Yeast killer plasmid pGKL1 encodes a DNA polymerase belonging to the
RT family B DNA polymerases.";
RL Nucleic Acids Res. 15:9088-9088(1987).
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + {DNA} (N).
CC -!- MISCELLANEOUS: This DNA polymerase requires a protein as a primer.
CC -!- MISCELLANEOUS: The presence of the two linear plasmids, termed
CC pGKL-1 and pGKL-2, in strains of Kluyveromyces lactis confers the
CC killer phenotype, i.e., production of toxin and resistance to it,
CC to the host cell.
CC -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X07127; CAA30136.1; ALT_INIT.
DR

RT "Human apolipoprotein B: structure of carboxyl-terminal domains,
RT sites of gene expression, and chromosomal localization.";
RL Science 230:37-43(1995).
RN [10]
RP SEQUENCE OF 1-291 FROM N.A.
RX MEDLINE=86149325; PubMed=3513177;
RA Protter A.A., Hardman D.A., Schilling J.W., Miller J., Appleby V.,
RA Chen G.C., Kirsner S.W., McEnroe G., Kane J.P.;
RT "Isolation of a cDNA clone encoding the amino-terminal region of
RT human apolipoprotein B.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:1467-1471(1986).
RN [11]
RP SEQUENCE OF 1-1670 FROM N.A.
RX MEDLINE=86287319; PubMed=3461454;
RA Protter A.A., Hardman D.A., Sato K.Y., Schilling J.W.,
RA Yamanaka M., Hort Y.J., Hjerrild K.A., Chen G.C., Kane J.P.;
RT "Analysis of cDNA clones encoding the entire B-26 region of human
RT apolipoprotein B.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:5678-5682(1986).
RN [12]
RP PARTIAL SEQUENCE, AND IDENTIFICATION OF APO-B48.
RX MEDLINE=88018019; PubMed=3659919;
RA Chen S.-H., Habib G., Yang C.-H., Gu Z.-W., Lee B.R., Weng S.-H.,
RA Silberman S.R., Cai S.-J., Deshpere J.P., Rosseneu M.,
RA Gotto A.M. Jr., Li W.-H., Chan L.;
RT "Apolipoprotein B-48 is the product of a messenger RNA with an organ-
RT specific in-frame stop codon.";
RL Science 238:363-366(1987).
RN [13]
RP DOMAINS.
RX MEDLINE=87039351; PubMed=3773997;
RA Knott T.C., Pease R.J., Powell L.M., Wallis S.C., Rall S.C. Jr.,
RA Innerarity T.L., Blackhart B., Taylor W.R., Marcel Y., Milne R.,
RA Johnson D., Fuller M., Lusis A.J., McCarthy B.J., Mahley R.W.,
RA Levy-Wilson B., Scott J.;
RT "Complete protein sequence and identification of structural domains
RT of human apolipoprotein B.";
RL Nature 323:734-738(1986).
RN [14]
RP DOMAINS.
RA Yang C.-Y., Chen S.-H., Gianturco S.H., Bradley W.A., Sparrow J.T.,
RA Tanimura M., Li W.-H., Sparrow D.A., Deloof H., Rosseneu M.,
RA Lee F.-S., Gu Z.-W., Gotto A.M. Jr., Chan L.;
RT "Sequence, structure, receptor-binding domains and internal repeats
RT of human apolipoprotein B-100.";
RL Nature 323:738-742(1986).
RN [15]
RP CALCIUM-BINDING DATA.
RX MEDLINE=86242245; PubMed=3087360;
RA Dashti N., Lee D.M., Mok T.;
RT "Apolipoprotein B is a calcium binding protein.";
RL Biochem. Biophys. Res. Commun. 137:493-499(1986).
RN [16]
RP PALMITOYLATION OF CYS-1112.
RX MEDLINE=20143590; PubMed=10679026;
RA Zhao Y., McCabe J.B., Vance J., Berthiaume L.G.;
RT "Palmitoylation of apolipoprotein B is required for proper
RT intracellular sorting and transport of cholesterol esters and
RT triglycerides.";
RL Mol. Biol. Cell 11:721-734(2000).
RN [17]
RP VARIANT SER-4338.
RX MEDLINE=91071750; PubMed=1979313;
RA Nava'as M., Laurent A.-M., Moreel J.-F., Ragab A., Cambou J.-P.,
RA Cuny G., Cambien F., Roizes G.;
RT "Detection by denaturing gradient gel electrophoresis of a new
RT polymorphism in the apolipoprotein B gene.";
RL Hum. Genet. 86:91-93(1990).
RN [18]
RP VARIANT FDB GLN-3527.
RX MEDLINE=89098975; PubMed=2563166;
RA Soria L.F., Ludwig E.H., Clarke H.R.G., Vega G.L., Grundy S.M.,
RA McCarthy B.J.;

RT "Association between a specific apolipoprotein B mutation and
RT familial defective apolipoprotein B-100.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:587-591(1989).
RN [19]
RP VARIANT LEU-2739.
RX MEDLINE=91016974; PubMed=2216805;
RA Huang L.-S., Gavish D., Breslow J.L.;
RT "Sequence polymorphism in the human apoB gene at position 8344.";
RL Nucleic Acids Res. 18:5922-5922(1990).
RN [20]
RP VARIANT FDB CYS-3558.
RX MEDLINE=95190020; PubMed=7883971;
RA Pullinger C.R., Hennessy L.K., Chatterton J.E., Liu W., Love J.A.,
RA Mendel C.M., Frost P.H., Malloy M.J., Schumaker V.N., Kane J.P.;
RT "Familial ligand-defective apolipoprotein B. Identification of a new
RT mutation that decreases LDL receptor binding affinity.";
RL J. Clin. Invest. 95:1225-1234(1995).
RN [21]
RP VARIANTS LEU-1437; SER-1914; LYS-2566; THR-3121; ALA-3945; MET-4128
RP AND THR-4481.
RX MEDLINE=97044521; PubMed=8889592;
RA Poirier O., Ricard S., Behague I., Souriau C., Evans A.E.,
RA Arveiler D., Marques-Vidal P., Luc G., Roizes G., Cambien F.;
RT "Detection of new variants in the apolipoprotein B (Apo B) gene by
RT PCR-SSCP.";
RL Hum. Mutat. 8:282-285(1996).
RN [22]
RP VARIANTS FDB GLN-3527 AND CYS-3558.
RX MEDLINE=97403938; PubMed=9259199;
RA Rabes J.P., Varret M., Saint-Jore B., Erlich D., Jondeau G.,
RA Krempf M., Giraudet P., Junien C., Boileau C.;
RT "Familial ligand-defective apolipoprotein B-100: simultaneous
RT detection of the ARG3500->GLN and ARG3531->CYS mutations in a
RT French population.";
RL Hum. Mutat. 10:160-163(1997).
RN [23]
RP VARIANTS SER-1914; ARG-1923; LEU-2739; ASP-3319; THR-3427; GLN-3432
RP AND ILE-3921.
RX MEDLINE=98141125; PubMed=9490296;
RA Lerer T.P., Bakken K.S., Hoel V., Hjermann I., Berg K.;
RT "Screening for mutations of the apolipoprotein B gene causing
RT hypcholesterolemia.";
RL Hum. Genet. 102:44-49(1998).
CC -!- FUNCTION: APOLOPROTEIN B IS A MAJOR PROTEIN CONSTITUENT OF
CC CHYLOMICRONS, VLDL AND LDL. IT FUNCTIONS AS A RECOGNITION SIGNAL
CC FOR THE CELLULAR BINDING AND INTERNALIZATION OF LDL PARTICLES BY
CC THE APOB/E RECEPTOR.

Query Match 3.3%; Score 116; DB 1; Length 4563;
Best Local Similarity 16.3%; Pred. No. 29;
Matches 104; Conservative 94; Mismatches 197; Indels 244; Gaps 26;
QY 218 MEFLNSY-----HQADY-----LSTHYAEQGNHRLFEA-----QRNL----- 250
Db 3068 IDFLNLYALFLSPSQASQVARSARPNKYQNFSAGNNIMEAHVINGEANDLFLN 3127
QY 251 -----FAGVSFFPEFKDSPRWQTG----- 269
Db 3128 IPLTIPEMELPYTIITPLKDFSLWEKTLTKQSFDSLVSVAQYKKNKHSIT 3187
QY 270 ---ISVUNTEIKQVADGMQFELSPYHVAADIFLKAYSAGK-----RVNLEK---EFFQ 320
Db 3188 NFLAVLCEPISQSIKSFDRHFENK-----RNALDFVTKSYNETKIKPDKYKAKSHDELPR 3244
QY 321 SVYQVTVENMIMALISISLPDYNTPMGDSWITDKNFRMAQF-----ASMARVFPANQAIKY 376
Db 3245 TP-----QIEGYTVPVV-----NVEVSPFTIEMSAFYGVFPKAVSMPS 3282
QY 377 FATDGKQKGKAPNF-----LSKALSNAGF-YT 401
Db 3283 FSIIGSDVRVPSYTLPLSLPLVHLVPRNLKLSLPHFKELCTISHIFIPAMGNITYDFS 3342
QY 402 FRSGWGDKNATVWLKASPFGEFHAQPDNGTFFELFIKGRNFTPDAGVVFVYSGDEAI----- 456


```
FT VARIANT 1050 1050 L -> F (IN STRAIN PS178D4).
FT CONFLICT 1094 1118 GRHQTACTWKRONSVRSWTKR -> EDEYKLYVRGK
SQ SEQUENCE 1167 AA; 131658 MW; 47ALED4EBDC645 CRC64;

Query Match
Best Local Similarity 3.3%; Score 114.5; DB 1; Length 1167;
Matches 103; Conservative 62; Mismatches 170; Indels 189; Gaps 24;

QY 249 NLFAGVSPPEF--KDSRWRQTGIVLNTKQV-----YADGQFELS 291
DQ ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 94 NMFVDIFPKLFGSGSQNAQAFELIEKVELVDEFRNFTLNLLNLYLDGQTALS 153
QY 292 PIYHVAADIFLKAGSKRVNLEKEFPQSVQTVENMI-----MALISILPDYN 342
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 154 HFQNDVQIAI---COGEPGLMD--QTACTPTTDHLISVRESFKDARTIETALPHK 209
QY 343 TPFMGDSWITDKNFRMAQFASWARVFPANQAIKFATDGKQKAFNLSKALS---AG 398
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 210 NPMLS-----TNDNTPDFNSDVLTLTPWYTTGATLNLILHQQ 247
QY 399 FYTPRSGWDK-----NATVWLKASPPGEFHAQPDNGTFELIKGNFTPDAGVFY 450
DB : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 248 YIQFAERKSVNYDESPINQTKVOLRRI--QDYSTTVSTTPEKFKPTLN-----P 296
QY 451 SGDEAIMKLRWYQTRIHST-----LTLDN-----QNMVITKARQKWE-- 490
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 297 SNKESVKNYKRVSRMTLQSLDIAATWPTLDNVNVPNSVDIQLDQTRLVFSDVAGPWE 356
QY 491 ---TCNNLDVLTYPN-----SYNLDHQ-----RSV-----LF 516
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 357 DNITSNIIDVLTPIGTIGFQESSDLRKFTYPIELQSMQPHGQVNSKSVCHYSDGLK 416
QY 517 INKKYFLVIDRAIGATGNLGVHWQKEDSN-----PVFDKTRVYTYRD 563
DB : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 417 LNYK-----NKTITAGVSN-----DESNQNNKHNGPVINSPITDINVNSQSLD 464
QY 564 GNNLMIQ-----SLNADRTSLNEEKGKSVYVYNKELKRPAPFEPKKNAGTQ- 611
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 465 LNSVMVNGGQKVTGCSPLSSNGSNNAALPNQKINVISVQSN-----DKPEKHADIYR 518
QY 612 ---NFVSIYVPYD-----GQKAPETISIRENKGNDFPKG 641
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 519 KWGYMSSHIPYDLVPENVIGDIDPTDKPSLLL---KGPPAEKG 559

RESULT 10
EX5B_HAEIN STANDARD; PRT; 1211 AA.
AC P45157;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Exodeoxyribonuclease v beta chain (EC 3.1.11.5).
GN RECB OR H1321.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RC MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Keiley J.M.,
RA Weidman J.P., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrman J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae"
```

```
RT Rd.";
RL Science 269:496-512 (1995).
CC !- FUNCTION: REQUIRED FOR EFFICIENT DNA REPAIR; IT CATALYZES THE
CC UNWINDING OF DOUBLE-STRANDED DNA AND THE CLEAVAGE OF SINGLE-
CC STRANDED DNA AND IT STIMULATES LOCAL GENETIC RECOMBINATION.
CC ALL OF THESE ACTIVITIES REQUIRE CONCOMITANT HYDROLYSIS OF ATP
CC (BY SIMILARITY).
CC !- CATALYTIC ACTIVITY: Exonucleolytic cleavage (in the presence of
CC ATP) in either 5'- to 3'- or 3'- to 5'-direction to yield 5'-
CC phosphooligonucleotides.
CC !- SUBUNIT: CONSIST OF THREE SUBUNITS; RECB, RECC AND RECD
CC (BY SIMILARITY).
CC !- SIMILARITY: BELONGS TO THE HELICASE FAMILY. UVRD SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32811; AAC22966.1; -.
DR PIR; D64116; D64116.
DR HSSP; P09980; 1UAA.
DR TIGR; H11321; -.
DR InterPro; IPR004586; RecB.
DR InterPro; IPR000212; UvrD-helicase.
DR Pfam; PF00580; UvrD-helicase; 1.
DR TIGRFAMs; TIGR00609; recB; 1.
KW Hydrolase; Nuclease; Exonuclease; Endonuclease; Helicase; ATP-binding;
KW DNA repair; Complete proteome.
FT NP BIND 22 29 ATP (POTENTIAL)
SQ SEQUENCE 1211 AA; 139857 MW; 37B76430651PAD68 CRC64;

Query Match
Best Local Similarity 3.3%; Score 114.5; DB 1; Length 1211;
Matches 102; Conservative 67; Mismatches 190; Indels 149; Gaps 23;

QY 53 AGNYDDAAKALLAYVREKSKAREPDSNAEKPADIR---QPDKVTRMADKALVHQP- 107
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 725 AATLHSEAAALLSWFEKQIQ-----GEARQEQIRLESEKQLVKIVTIHKSGLGYDILV 778
QY 108 -----OPHKGYGFDYCKOINWQMPVKD-----NEVRW---QLHRVK 142
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 779 WLPFLAAPSCKPSKXYINIIYSKERDETLNDIENLNALCEETFAEELRLLYVALTRAK 838
QY 143 WQQAALVYHATGDEKY-AREWVYQYSDWARKNPLGLSQDNNDKFWVRPLE-VSDRVQ- 197
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 839 YQMAFAL--PAQFDKKNALHYVLVSQGEIKELNLSKDTETLLQTFEKKMQDNVEICT 896
QY 198 ----SLPTFSLFVNSPAPTAFLMEFLNSVHQOADYLSHYAEQGNHRLFEAQNLIF-- 251
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 897 KPNLEALPTLSINTKNDDFKAS---EFTGNIEQDWRTSFTSIEQAHR- ---QNYFTE 948
QY 252 -AGVSFPPEFKOS-----PRMRTGTIS-----V 272
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 949 SAGKKHAFDDAKDYDSQNAIEISTALLNENESNILDIPRGKQVGTALHRRHFCYFSDL 1008
QY 273 LNT-BIKK-----QVYADGMQFELSPIYHVAIDIFLKAYGSAKRVNLEKEFPQSV 323
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1009 ANTEBIDKRLQSLQDDETFSTESLQNLQWQISHTPLSNEIGIALADLANCKDCKEMP-FYL 1067
QY 324 QTVENMIMALISILPDYNTVMTFGDSWITDKNFRMAQFASWAR-----VFPANQAIKFA 378
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1068 AIREH-----FDVEARNHILKAHHLPSLSLOFEQIQGMVRSIDIVFRHNG--KYL 1118
QY 379 TDGKQKAPNLFSLKALSNA-----FYTPRSGWKNATVWLKASPP 420
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1119 VDYKS---NFLGSLADVNOEALKKEMLSHYDWOYLIYTL-----ALHRYLSQSVVP 1167
QY 421 GEFHAQPDNGTFELIKGNFTPDAGVF 448
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```


SIGNAL	1	?	POTENTIAL
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			
25			
26			
27			
28			
29			
30			
31			
32			
33			
34			
35			
36			
37			
38			
39			
40			
41			
42			
43			
44			
45			
46			
47			
48			
49			
50			
51			
52			
53			
54			
55			
56			
57			
58			
59			
60			
61			
62			
63			
64			
65			
66			
67			
68			
69			
70			
71			
72			
73			
74			
75			
76			
77			
78			
79			
80			
81			
82			
83			
84			
85			
86			
87			
88			
89			
90			
91			
92			
93			
94			
95			
96			
97			
98			
99			
100			

QY 548 PVFDKTKNRVY-----TYRDGNNLMIQSLNADRTSLNEEGKVSYY----- 590
Dd 1114 -----EARNYTLIVKHALTEPVNAIV--LIIGNYTYLTDBNGVTFYTAPTKGSDEI 1165
QY 591 -----NKLKRPAPVFEKPKKNAQTQNFVSIVPYDG-- 622
Dd 1166 TVIVKKNFNTLEKTFQITVSEPEITEEDINEPKLMSPEANA---TIVVEMSESGV 1222
QY 623 QKAPISIRENGDNFEKGLNLTITNGKQQLVAVP 659
Dd 1223 KKTIVTEITIN-----GTANETATI-----WVVP 1247

RESULT 12

POLG_EC23W STANDARD; PRT; 2179 AA.
AC 073556;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polypeptide [Contains: Coat protein VP0 (PIAB); Coat protein VP3 (PIC); Coat protein VP1 (PID); Core protein 2A; Core protein P2B; Core protein P2C; Core protein P3A; Genome-linked protein VP0 (P3B); Picornain 3C (EC 3.4.22.28) (Protease 3C) (P3C); RNA-directed RNA polymerase (EC 2.7.7.48) (P3D)].
DE Echovirus 23 (strain Williamson) (Human parechovirus 2).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae; Parechovirus.
OC NCBI TaxID=122962;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99036025; PubMed=9820139;
RA Ghazi F., Hughes P.J., Hyypiae T., Stanway G.;
RT "Molecular analysis of human parechovirus 2 (echovirus 23).";
RL J. Gen. Virol. 79:2641-2650(1998).
CC -!- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN Q/G SITES IN THE POLYPEPTIDE. IT IS A CYSTEINE PROTEASE.
CC -!- CATALYTIC ACTIVITY: Selective cleavage of Gln-Gly bond in the poliovirus polypeptide. In other picornavirus reactions Glu may be substituted for Gln, and Ser or Thr for Gly.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + {RNA} (N).
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS, EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP0, VP1, AND VP3.
CC -!- PTM: SPECIFIC ENZYMIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -!- SIMILARITY: P3C PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AJ005695; CAA06679.1; -.
DR MEROPS; C03.023; -.
DR InterPro; IPR007053; NC.
DR InterPro; IPR000605; RNA_helicase.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR001205; RNA_pol_P3D.
DR InterPro; IPR007094; RNA_pol_Psvir.
DR Pfam; PF04970; NC; 1.
DR Pfam; PF00680; RNA_dep_RNA_pol; 1.
DR Pfam; PF00910; RNA_helicase; 1.
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT CHAIN 1 289 COAT PROTEIN VP0.
FT CHAIN 290 542 COAT PROTEIN VP3.
FT CHAIN 543 775 COAT PROTEIN VP1.

FT CHAIN 776 922 CORE PROTEIN P2A.
FT CHAIN 923 1044 CORE PROTEIN P2B.
FT CHAIN 1045 1373 CORE PROTEIN P2C.
FT CHAIN 1374 1490 CORE PROTEIN P3A.
FT CHAIN 1491 1510 GENOME-LINKED PROTEIN VP0.
FT CHAIN 1511 1710 PICORNAIN 3C.
FT CHAIN 1711 2179 RNA-DIRECTED RNA POLYMERASE.
FT SITE 763 765 CELL ATTACHMENT SITE (POTENTIAL).
FT ACT_SITE 1669 1669 PROTEASE 3C (POTENTIAL).
FT ACT_SITE 1687 1687 PROTEASE 3C (POTENTIAL).
SQ SEQUENCE 2179 AA; 245872 MW; 96803COBB8856664 CRC64;

Query Match

Best Local Similarity 3.2%; Score 113; DB 1; Length 2179;

Matches 140; Conservative 89; Mismatches 261; Indels 210; Gaps 36;

QY 4 KIFKRIIVFAVIALSSGNIQAQSSITRKDFDHINLEYSGLKYNKAVAGNDYDAAKAL 63
Dd 953 KILVRILCYMVLVCHKNILTTACLSTLLVMDVTS--SSVLSPSCALMQLMDGDKL 1010
QY 64 LAYVEKSKAREPDFSNAEKPADIRQPDVKVTREMA-----KALVHQFQPHKGYGFDY 118
Dd 1011 AEVVAE-----SMSNTDD-----DEIKEQICDVTYTKQILSNQGPFGF----- 1050
QY 119 GKIDINQMWPVKDNEVRWQLHRVKW-----QAMALVYHATGDEKYAREVWYQSDWA 171
Dd 1051 -----NEISTAFRHIDMWIQTLLKIKDWLVSFKPSVEKRAVEL----- 1090
QY 172 RKNPLGLSQDNDKFWRPLEVSDRVQSLPPTFSLFVNSPAPFAELMEFLNSYHQOAYL 231
Dd 1091 -----ERNKEHVCISILDYAS-----DIIVKSKQDTKMKTKQEFYQRYN---DCL 1130
QY 232 STHYAEQGNHRLFEAQRNLFAGVSPFEPKDSR-----WRQ-----T 268
Dd 1131 SKFKEIMA--MCFRCHNSISNTVYRLPQELARIPNRMATQNDLIRVEPIGIWIOGEPQ 1188
QY 269 GISVLNTRIKQOVAD-GMQFELSPIYHVAIDIFLKAYGSAKRVNLEKFFQSVQVTV 327
Dd 1189 GKSFLTHLTKQLQKTCGLQ-----GIYTNPTASEFMDGYDN-QDIHLIDDLGQTRKERDI 1243
QY 328 NMIMALISLSLDYNTFPMFGSDWITDKFRMAQFASWARVFAQNAIKYFATDGGKAP 387
Dd 1244 EMLCNCIS--SDPDI-----VPMALHEKKGKFTYSKLVI---ATTNK-----P 1281
QY 388 NFLSKALSNAAGFYTFERSGMDKNATVMVLA-----SPPGEFH---AQPDMGTFFELFKGRN 440
Dd 1282 DFFSVTLVDSGALRRFPY-----IMHFEAAHYSKSKLVNSQAMPHMSTGCEWVSKN 1336
QY 441 FTFDAGVYVYSGDEAIMKLRNRYOTRIHSTLTLDNQNNVITKARQNW-----ETGN 494
Dd 1337 -----GRDWETLKLK-----ELIDKITVDYKERI---ANYNTWKQLDQTLDD 1377
QY 495 L-DVLYTNPSPN-----LDHQRSVLFINKKYFL-----VIDRAIGEATGNLGVHWQ 541
Dd 1378 LDDAVSYTKHNPYDAIPYIDEVLIENIEMSTLIEQMEAFIEPKPSVFKCFASRVGD-----K 1432
QY 542 LKEDSNPFV-----DKTKNRVYTYRDGNNLMQSLNADRTSL-----NEEGKVS 587
Dd 1433 IKAESREVVKWFSKDKLSML--NFVERNKAMLTVVSVAITSAIGILLVTKIFKKEESKDE 1490
QY 588 YVYNKELKRPAPVFEKPKKNAQT-----QNFVSIVPYDGQ 623
Dd 1491 RAYN-----PTLPVAKPK---GTFPVFSQREFKNEAPYDQG 1522

RESULT 13

INVL_YEAST STANDARD; PRT; 532 AA.
ID INVL_YEAST
AC P10594;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Invertase 1 precursor (EC 3.2.1.26) (Beta-fructofuranosidase 1)

```

DE (Saccharase).
GN SUC1.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE OF 1-73 FROM N.A.
RX MEDLINE=88216256; PubMed=2835632;
RA Hohmann S.; Gozalbo D.;
RT "Structural analysis of the 5' regions of yeast SUC genes revealed
RT analogous palindromes in SUC, MAL and GAL.";
RL Mol. Gen. Genet. 211:446-454(1988).
RN [2]
RP SEQUENCE FROM N.A.
RA Hohmann S.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-
CC fructofuranoside residues in beta-D-fructofuranosides.
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X07570; CAA30457.1; -
CC PIR; S27372; S27372.
CC SCD; L0002143; SUC1.
CC InterPro: IPR001362; Glyco_hydro_32.
CC Pfam: PF00251; Glyco_hydro_32; 1.
CC SMART; SM00640; Glyco_32; 1.
CC PROSITE; PS00609; GLYCOSYL HYDROL F32; 1.
CC Hydrolase; Glycosidase; Glycoprotein; Multigene family; Signal.
FT SIGNAL 1 19
FT CHAIN 20 532
FT ACT_SITE 42 42
FT CARBOHYD 23 23
FT CARBOHYD 64 64
FT CARBOHYD 111 111
FT CARBOHYD 112 112
FT CARBOHYD 118 118
FT CARBOHYD 165 165
FT CARBOHYD 275 275
FT CARBOHYD 356 356
FT CARBOHYD 369 369
FT CARBOHYD 384 384
FT CARBOHYD 398 398
FT CARBOHYD 512 512
FT SEQUENCE 532 AA; 60570 MW; 3280EF260EA01606 CRC64;
Query Match 3.2%; Score 112; DB 1; Length 532;
Best Local Similarity 18.6%; Pred. No. 2.6;
Matches 107; Conservative 83; Mismatches 176; Indels 210; Gaps 31;
QY 94 VTREMAKALVHQFQPHKGY-----GYFDYKIDINQMW-PVKDNEVRQLHRVQWQAM 147
DB 21 MTNETSDRPLVH-FTFNKGWMDPNGLWYDAKEGKWHLYFYQNPNDTWGL--PLFWG-- 75
QY 148 ALVYHATGDKYAREWVYQSDWARKNPLGLSQDNDKFWWRPLEVSDRVQSFPPTFSLV 207
DB 76 ----HATSDD-----LTHW-----QD-----EPVAIAPRK-- 97
QY 208 NSPAPFAPLMBF-----LNSVHQADYLSTHYAEOGNHRLFE 245
DB 98 DSGAYSGSNVIDYNNSTGFFNDTIDPRQCVAIWYNTPESEQVLS--YSLDGGYTFTE 155
QY 246 AQRNLFAGVSPFEPKD-----SPRWRQTGISVLNTEIKQVYAD----- 284
DB 156 YQKNPVLAA NSTQFRDPKVFVWPEPKWTMT--AAKSQDYKIEIYSSDDLKSKWLESAPA 213

```

(Saccharase).

SUC1.

Saccharomyces cerevisiae (Baker's yeast).

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetales; Saccharomycetaceae; Saccharomyces.

NCBI_TaxID=4932;

[1]

SEQUENCE OF 1-73 FROM N.A.

MEDLINE=88216256; PubMed=2835632;

Hohmann S.; Gozalbo D.;

"Structural analysis of the 5' regions of yeast SUC genes revealed

analogous palindromes in SUC, MAL and GAL.";

Mol. Gen. Genet. 211:446-454(1988).

[2]

SEQUENCE FROM N.A.

Hohmann S.;

Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.

-!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-

fructofuranoside residues in beta-D-fructofuranosides.

-!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See http://www.isb-sib.ch/announce/

or send an email to license@isb-sib.ch).

EMBL; X07570; CAA30457.1; -

PIR; S27372; S27372.

SCD; L0002143; SUC1.

InterPro: IPR001362; Glyco_hydro_32.

Pfam: PF00251; Glyco_hydro_32; 1.

SMART; SM00640; Glyco_32; 1.

PROSITE; PS00609; GLYCOSYL HYDROL F32; 1.

Hydrolase; Glycosidase; Glycoprotein; Multigene family; Signal.

SIGNAL 1 19

CHAIN 20 532

ACT_SITE 42 42

CARBOHYD 23 23

CARBOHYD 64 64

CARBOHYD 111 111

CARBOHYD 112 112

CARBOHYD 118 118

CARBOHYD 165 165

CARBOHYD 275 275

CARBOHYD 356 356

CARBOHYD 369 369

CARBOHYD 384 384

CARBOHYD 398 398

CARBOHYD 512 512

SEQUENCE 532 AA; 60570 MW; 3280EF260EA01606 CRC64;

Query Match 3.2%; Score 112; DB 1; Length 532;

Best Local Similarity 18.6%; Pred. No. 2.6;

Matches 107; Conservative 83; Mismatches 176; Indels 210; Gaps 31;

QY 94 VTREMAKALVHQFQPHKGY-----GYFDYKIDINQMW-PVKDNEVRQLHRVQWQAM 147

DB 21 MTNETSDRPLVH-FTFNKGWMDPNGLWYDAKEGKWHLYFYQNPNDTWGL--PLFWG-- 75

QY 148 ALVYHATGDKYAREWVYQSDWARKNPLGLSQDNDKFWWRPLEVSDRVQSFPPTFSLV 207

DB 76 ----HATSDD-----LTHW-----QD-----EPVAIAPRK-- 97

QY 208 NSPAPFAPLMBF-----LNSVHQADYLSTHYAEOGNHRLFE 245

DB 98 DSGAYSGSNVIDYNNSTGFFNDTIDPRQCVAIWYNTPESEQVLS--YSLDGGYTFTE 155

QY 246 AQRNLFAGVSPFEPKD-----SPRWRQTGISVLNTEIKQVYAD----- 284

DB 156 YQKNPVLAA NSTQFRDPKVFVWPEPKWTMT--AAKSQDYKIEIYSSDDLKSKWLESAPA 213

QY 285 -----GMQPEL-----SPIYHVAADI-----FLKAY 306

DB 214 NEGFGLGYQYECPLIEVPSEQDFSKSHWYMFISINFGAPAGGSFNQYFVGSFNGHHFEAF 273

QY 307 GSAKRVNLEKEFPQSY--VQTVENMIMALISLSLPDYNTPMFGDSWITDKNPEMAOP--- 361

DB 274 DNQSRV---VDFGKDYVALQTFN-----TDPYGSAL-GIAWAS--NWEYSAPVPS 319

QY 362 ----ASWARVFPANQAIKYFA---TDGKQKAPNLSKALSNAAGFYTFERSGWDKNATVMV 414

DB 320 NPWRSSMSLVRFPSLNTYQANPEFELINLKAEPILN--ISSAG-----PWSRFATNTT 371

QY 415 LKASPPGEFHAQPDNGTFELFIKGRNFTPDAGVFVYSGDEAIMK-----LRNWRQTR-- 467

DB 372 LTKANSYNDLSNSTGTLEFEL-----VYAVNTTQTISKSVFADLSLWFKGLEDP 421

QY 468 -----IHSTLTDNONNVITKARONKWEFG---NNLDVLTYTNPSYPN---LD 509

DB 422 EBYLRMGFEVSASSFELDRGNSKVFEKENPFTVRMSVNNQFPKSENDSLYYKVGILLD 481

QY 510 HORSVLFIN-----KKYFLVIDRAIGBATCNLGV 538

DB 482 QNILELYFNDGDWSTNTYEMTTGNALGSVNMVTGV 517

RESULT 14

INV4_YEAST

ID INV4_YEAST STANDARD; PRT; 532 AA.

AC P10596;

DT 01-JUL-1989 (Rel. 11, Created)

DT 01-JUL-1989 (Rel. 11, Last sequence update)

DT 01-MAR-1992 (Rel. 21, Last annotation update)

DE Invertase 4 precursor (EC 3.2.1.26) (Beta-fructofuranosidase 4)

DE (Saccharase).

OS SUC4.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.

OX NCBI_TaxID=4932;

RN [1]

RP SEQUENCE OF 1-73 FROM N.A.

RX MEDLINE=88216256; PubMed=2835632;

RA Hohmann S.; Gozalbo D.;

RT "Structural analysis of the 5' regions of yeast SUC genes revealed

analogous palindromes in SUC, MAL and GAL.";

RL Mol. Gen. Genet. 211:446-454(1988).

RN [2]

RP SEQUENCE FROM N.A.

RA Hohmann S.;

RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.

CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-

fructofuranoside residues in beta-D-fructofuranosides.

CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See http://www.isb-sib.ch/announce/

or send an email to license@isb-sib.ch).

CC -----

CC EMBL; X07572; CAA30459.1; -

DR PIR; S27373; S27373.

DR SGD; L0002146; SUC4.

DR InterPro: IPR001362; Glyco_hydro_32.

DR Pfam; PF00251; Glyco_hydro_32; 1.

DR SMART; SM00640; Glyco_32; 1.

DR PROSITE; PS00609; GLYCOSYL HYDROL F32; 1.

KW Hydrolase; Glycosidase; Glycoprotein; Multigene family; Signal.

FT SIGNAL 1 19

FT CHAIN 20 532

FT INVERTASE 4.


```

QY 411 TVMVLKASPPGEFHAQPDNGTFELFIKGRNFTPDAGVFVYSGDEAIMKLENWYQTRIHS 470
Db 550 RVRVR-----YATNVSCQFNVI-----567
QY 471 TLTLDNONNVITKARONKWET-GNNLDVLTNTNPSYPNLDHQRSVLFINKKYFLVIDRAI 529
Db 568 -----NDKITLQRFQNTVETIGEGKD-LTYGSFGY--IEYSTTIQFPDKHPKITLH--L 617
QY 530 GEATGNLGVHWQLKE-----DSNPVFDKXNRVYTYTYRDGNNLM---IQSLNADRTS 578
Db 618 SDSLNNSSFYVDSIEFIPVDVNVYDEKEKLEKAQKAVNTLFTEGRNALQKDVTDYKVDQVS 677
QY 579 L 579
Db 678 I 678

```

Search completed: July 28, 2003, 20:50:14
Job time : 31 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 28, 2003, 20:46:28 ; Search time 46 seconds
(without alignments)
1377.721 Million cell updates/sec

Title: US-09-802-285A-2

Perfect score: 3494

Sequence: 1 MTTKIFKRIIVFAVIALSSG.....KGKLNLTITNGKQLVLVP 659

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 76:*

1: piri:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3443	98.5	659	JC4910	heparin-sulfate ly
2	189	5.4	633	A97309	hypothetical prote
3	176	5.0	546	F95038	conserved hypothet
4	160	4.6	1183	F90559	poly(beta-D-mannur
5	154	4.4	713	JC5870	hypothetical prote
6	139	4.0	561	T36084	p60-related protei
7	139	4.0	764	AB1695	hypothetical prote
8	137.5	3.9	604	T44835	hypothetical prote
9	127	3.6	560	E96554	hypothetical prote
10	125.5	3.6	672	AC0422	hypothetical prote
11	124.5	3.6	1883	G82875	beta-fructofuranos
12	123	3.5	532	IFBY	hypothetical prote
13	120.5	3.4	625	H90533	probable sulfatase
14	120.5	3.4	970	E70533	hypothetical prote
15	120	3.4	1272	C90593	lipoprotein (impor
16	119	3.4	776	E90591	levanase (EC 3.2.1
17	118.5	3.4	677	A27286	RF1 protein - yeas
18	117.5	3.4	995	S07353	toxin-like outer m
19	117.5	3.4	2902	C71953	fibronectin-bindin
20	116	3.3	1117	S33851	apolipoprotein B-1
21	116	3.3	4563	LPRUB	MGI40 homolog - My
22	115.5	3.3	1113	F73327	probable membrane
23	115.5	3.3	1314	S19488	hemoglobin-binding
24	115	3.3	953	B64083	alpha-amylase (EC
25	115	3.3	1256	G97902	alkaline amylopul
26	115	3.3	1280	E95031	probable flagellin
27	114.5	3.3	750	G81361	exodeoxyribonuclea
28	114.5	3.3	1211	D64116	type IIS restricti
29	114.5	3.3	1279	E64709	

ALIGNMENTS

RESULT 1

JC4910

heparin-sulfate lyase (EC 4.2.2.8) - Flavobacterium heparinum

N;Alternate names: heparin-sulfate eliminase; heparinase III

C;Species: Flavobacterium heparinum

C;Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 03-Jun-2002

C;Accession: JC4910

R;Godavarti, R.; Davis, M.; Venkataraman, G.; Cooney, C.; Langer, R.; Sasisekharan, R.

Biochem. Biophys. Res. Commun. 225, 751-758, 1996

A;Title: Heparinase III from Flavobacterium heparinum: Cloning and recombinant expression

A;Reference number: JC4910; MUID:96374394; PMID:8780685

A;Accession: JC4910

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-659 <GD>

A;Note: the authors translated the codon ACC for residue 313 as Asn

C;Comment: This enzyme cleaves heparan sulfate-rich regions of acidic polysaccharides.

C;Superfamily: Flavobacterium heparinum heparitin-sulfate lyase

C;Keywords: carbon-oxygen lyase

Query Match 98.5%; Score 3443; DB 2; Length 659;
Best Local Similarity 98.8%; Pred. No. 8.5e-237;
Matches 651; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY	1	MTTKIFKRIIVFAVIALSSGNI	LAQSSSITRKDFD	HINLEYSGLKVNKAVAGNYDDAA	60
Db	1	MTTKIFKRIIVFAVIALSSGNI	LAQSSSITRKDFD	HINLEYSGLKVNKAVAGNYDDAA	60
QY	61	KALLAYREKSKAREP	DFSNAEKPADIR	QIDKVTREMAADKALVHOFPHKGYGYDYGK	120
Db	61	KALLAYREKSKAREP	DFSNAEKPADIR	QIDKVTREMAADKALVHOFPHKGYGYDYGK	120
QY	121	DINWQWPKDNEVQWQLHRV	KWQAMALVYHATG	DEKYAREWVVOYSDWAKRNPLGLSQ	180
Db	121	DINWQWPKDNEVQWQLHRV	KWQAMALVYHATG	DEKYAREWVVOYSDWAKRNPLGLSQ	180
QY	181	DNDKFWVRPLEVSDRVQSL	PTPTFSLFVNSP	ATPAFLMEFLNSYHQADYLSTHYAEQGN	240
Db	181	DNDKFWVRPLEVSDRVQSL	PTPTFSLFVNSP	ATPAFLMEFLNSYHQADYLSTHYAEQGN	240
QY	241	HLFLFAQNLFAGVSPFPK	SPRWRQTGISVLN	TEIKKQVYADGQMFELSPIYHVAID	300
Db	241	HLFLFAQNLFAGVSPFPK	SPRWRQTGISVLN	TEIKKQVYADGQMFELSPIYHVAID	300
QY	301	IFLKAYSKAKRNLEKEP	QSVQVQVENMIAL	ISISLPDYNTPMGSWITDKNFRMAQ	360
Db	301	IFLKAYSKAKRNLEKEP	QSVQVQVENMIAL	ISISLPDYNTPMGSWITDKNFRMAQ	360
QY	361	FASWARVFPANQAIKYF	ATDGGKQKQKAPN	FLSKALSNAGFYTFRSQWKNATVWLKASPP	420
Db	361	FASWARVFPANQAIKYF	ATDGGKQKQKAPN	FLSKALSNAGFYTFRSQWKNATVWLKASPP	420

QY 421 GEFHAQPDNGTFELFIKGRNFTPDAGVYVSGDEAIMKLRNWRQTRIHSITLTDNQNV 480
Db |||||
QY 421 GEFHAQPDNGTFELFIKGRNFTPDAGVYVSGDEAIMKLRNWRQTRIHSITLTDNQNV 480
Db |||||
QY 481 ITKARQNKWETGNLNDVLTNPSPNLDHORSVLFINKKYFLVIDRAIGSATGNLGVHW 540
Db |||||
QY 541 QLKEDSNPFDKTRVVTYTRDGNLMIOSLNADRTSLNEBEGKSVYVYNKELKRPV 600
Db |||||
QY 541 QLKEDSNPFDKTRVVTYTRDGNLMIOSLNADRTSLNEBEGKSVYVYNKELKRPV 600
Db |||||
QY 601 FEKPKNAGTQNFVSIYVYDQKAPESIRENKNDPEKGLNLTITNGKQQLVLP 659
Db |||||
QY 601 FEKPKNAGTQNFVSIYVYDQKAPESIRENKNDPEKGLNLTITNGKQQLVLP 659
Db |||||

RESULT 2

A97909
hypothetical protein spr0297 [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence revision 22-Oct-2001 #text_change 22-Oct-2001
C:Accession: A97909
R:Hoekins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; H
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 193, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; PMID:21429245; PMID:11544234
A:Accession: A97909
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-633 <KUR>
A:Cross-references: GB:AE007317; PIN:AAK99101.1; PID:g15457850; GSPDB:GN00174
C:Genetics:
A:Gene: spr0297

Query Match 5.4%; Score 189; DB 2; Length 633;
Best Local Similarity 21.1%; Pred. No. 2e-05;
Matches 139; Conservative 100; Mismatches 273; Indels 148; Gaps 28;
QY 71 SKAREPDSNAEKPADIRQIDKVTREMAKALVHGFQPHKGYGFDYDKDINQMWPVK 130
Db |||||
QY 131 DNEVWQHLRWKQWQAMALVYHATGDEKYAR-----EWYQYSDWAKRNPLGLSQND 183
Db |||||
QY 73 DPEWYMLNRQBYLLQFMIGYLVGDKGYIQCKKFFLDWIEQVREFSPQSLMTRDGTG 132
Db |||||
QY 184 --KFVWRPLEVSDRVQSLPPTFSLFVNSPAFTPAFLMEFLNSVHQADYLSHYABQ--- 238
Db |||||
QY 133 IRSFTWLK-----LLLLKFDLLEKELEKILVLEKQIDFMKSYIRAKYTL 180
Db |||||
QY 239 GNHRLFEAQRNLFAGVSFPFKDSPRWRTGTSVLNTEIKQVYADGMQFELSPIYHV-- 296
Db |||||
QY 181 SNWGILQITPMLAIYHFFSDKMDLEAHYFASSELKQIETQLGDSGQFQESILYHVEV 240
Db |||||
QY 297 --AADIIFLKAGSAKRVNLEKEFPQSYVQTVENMIMALISLPIYNTPMFGDSWITDK 354
Db |||||
QY 241 YKALLDLCL-----LLPDLQDSYQELLEKNATYIQMTGLDGRTLAFGSDSTET 290
Db |||||
QY 355 NFRMAQFA-----SWARVFPAQAIKYFATDQKQKAPNLSKALSNA 397
Db |||||
QY 291 TEMPLSASVNLKEDLLNGLDVKVLLSILFLGREKVKLQEPERRAWQPK--SMIFEDS 348
Db |||||
QY 398 GFYTFRSQWMDKNAVWVILKASPPGFHAQPDNGTTFELFIKGRNFTPDAGVYVSGDEAIM 457
Db |||||
QY 349 GHVCIKD-----EHRVLFKNGPLGSAHSHSDNSFCLOQCGQIFIDAGRYSI----- 397
Db |||||
QY 458 KLRNWRQTRI-----HSTLTDNQNVITKARQ---NKWE-----TGNNLDVLT 499
Db |||||
QY 398 --REIYERYLLKSAWHSHTCIVDG-----KAPERITGSWEVEYYPHSLFCHHKEREGMH 449
Db |||||

QY 500 YTNPSY-----PNLD--HORSVLFINCKYFLVID--RAIGEATGNLGVHQLKEDSNPVFD 551
Db |||||
QY 450 YIEGAYSAEPDLPVLHKKRKLMLVEDVWLLVDDIRCOGQ-----HEALTQF---ILD 499
Db |||||
QY 552 KTKNRVYTYTRDG--NNLMIOSLNADRTSLNEBEGKSVYVYNKELKRPV---FKPK 604
Db |||||
QY 500 KD-----VTYQDGKINQLRLWS-----EVDFOLEDITISPKYN--ELERSKLTIKRQFFNQ 549
Db |||||
QY 605 KKN---ACTQNFVSI---VYFYDQKAPESIRENKNDPEKGLNLTITNGKQQLVLP 658
Db |||||
QY 550 MLDYTIHAHSEFIRHSVYQTDREVENALAFENKNDKTLKLLLEDIRVGEKLCIV 609
Db |||||

RESULT 3

F95038
hypothetical protein SP0327 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: F95038
R:Tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hei
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapf
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
A:Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.
A:Reference number: A95000; PMID:21357209; PMID:11463916
A:Accession: F95038
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-546 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK74503.1; PID:g14971800; GSPDB:GN00164; TIGR:SP
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP0327

Query Match 5.0%; Score 176; DB 2; Length 546;
Best Local Similarity 20.8%; Pred. No. 0.00014;
Matches 117; Conservative 82; Mismatches 231; Indels 132; Gaps 23;
QY 71 SKAREPDSNAEKPADIRQIDKVTREMAKALVHGFQPHKGYGFDYDKDINQMWPVK 130
Db |||||
QY 26 SKDYEKVKSLERLIDNRPMFDS--PDMPEPCSKIHQIQP-----VWVQVFD 72
Db |||||
QY 131 DNEVWQHLRWKQWQAMALVYHATGDEKYAR-----EWYQYSDWAKRNPLGLSQND 183
Db |||||
QY 73 DPEWYMLNRQBYLLQFMIGYLVGDKGYIQCKKFFLDWIEQVREFSPQSLMTRDGTG 132
Db |||||
QY 184 --KFVWRPLEVSDRVQSLPPTFSLFVNSPAFTPAFLMEFLNSVHQADYLSHYABQ--- 238
Db |||||
QY 133 IRSFTWLK-----LLLLKFDLLEKELEKILVLEKQIDFMKSYIRAKYTL 180
Db |||||
QY 239 GNHRLFEAQRNLFAGVSFPFKDSPRWRTGTSVLNTEIKQVYADGMQFELSPIYHV-- 296
Db |||||
QY 181 SNWGILQITPMLAIYHFFSDKMDLEAHYFASSELKQIETQLGDSGQFQESILYHVEV 240
Db |||||
QY 297 --AADIIFLKAGSAKRVNLEKEFPQSYVQTVENMIMALISLPIYNTPMFGDSWITDK 354
Db |||||
QY 241 YKALLDLCL-----LLPDLQDSYQELLEKNATYIQMTGLDGRTLAFGSDSTET 290
Db |||||
QY 355 NFRMAQFA-----SWARVFPAQAIKYFATDQKQKAPNLSKALSNA 397
Db |||||
QY 291 TEMPLSASVNLKEDLLNGLDVKVLLSILFLGREKVKLQEPERRAWQPK--SMIFEDS 348
Db |||||
QY 398 GFYTFRSQWMDKNAVWVILKASPPGFHAQPDNGTTFELFIKGRNFTPDAGVYVSGDEAIM 457
Db |||||
QY 349 GHVCIKD-----EHRVLFKNGPLGSAHSHSDNSFCLOQCGQIFIDAGRYSI----- 397
Db |||||
QY 458 KLRNWRQTRI-----HSTLTDNQNVITKARQ---NKWE-----TGNNLDVLT 499
Db |||||
QY 398 --REIYERYLLKSAWHSHTCIVDG-----KAPERITGSWEVEYYPHSLFCHHKEREGMH 449
Db |||||
QY 500 YTNPSY-----PNLD--HORSVLFINCKYFLVID--RAIGEATGNLGVHQLKEDSNPVFD 551
Db |||||


```

Db 450 YIBGAWSAEPDLPYLHRRKILMLVEDVLLVDIRCOGQ-----HEVLTQF---ILD 499
QY 552 KTKNRVYTYTRDG--NNLMIQS 571
Db 500 KD-----VTYQDGKINQLRLWS 516

RESULT 4
F90559
conserved hypothetical protein MYPVU_3820 [imported] - Mycoplasma pulmonis (strain UAB CT)
C:Species: Mycoplasma pulmonis
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C:Accession: F90559
R:Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Gallisson, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm
A:Reference number: A99512; MUID:21267165; PMID:11353084
A:Accession: F90559
A:Status: preliminary
A:Genetic code: SGC3
A:Molecule type: DNA
A:Residues: 1-1183 <KUR>
A:Cross-references: GB:AL445566; PID:gl4089796; PIDN:CACI3555.1; GSPDB:GN00153
A:Experimental source: strain UAB CTIP
C:Genetics:
A:Gene: MYPVU_3820
A:Genetic code: SGC3

Query Match 4.6%; Score 160; DB 2; Length 1183;
Best Local Similarity 18.7%; Pred. No. 0.0062;
Matches 159; Conservative 131; Mismatches 252; Indels 308; Gaps 44;

QY 7 KRIL---VPAVIAL-----SSGNILAQSSITR-----KDFD----- 35
Db 4 KXIIASSFLPMSLLPFLSNTSTIAQEKXNITQVFKKNSLENFDEYRKNNDIEIGBKTL 63
QY 36 -----HINLEYGLEKVKVAAAGNVDDAAKALLAVYREKSKAREPDFSNAEKPAD 86
Db 64 IVSKFIETKLNLYKSDFEKINIVINDSNE-----NFSLGNNNSQ 105
QY 87 IRQPIDKVTREMAKALVHQFQPHKGYGPDYGDINQWMPVKDNEVRWQLHRVKWQA 146
Db 106 IE-----NWTNLNVKTNDDID-KLNSVKFF-- 128
QY 147 MALVYHATGDEKARWVYQYSDWARKNPLGLSQDNKDFWVRPLEVSDRVQSLPPTFSLF 206
Db 129 -----FNDKSSQSW-----DFIEITNIKQRIEKIKSDLSQK 160
QY 207 VNSPAPTEA-----FLME-----FLNSYHQADYL-----STHYAEOGNHRLFEAQ 247
Db 161 YNDNHESVSIKKLDFIDYKSPKNDLSFLNEFKKVDVILEIKRVNPFYKKEINSQIKIV 220
QY 248 RNLFAGVSPFEKDSR--WRQ-TG:SVLNLTEIKKQVYA---DGMQFELSPIYHVAIDI 301
Db 221 KD-FDKLNEFEFTSSLLFEKENVNLDDLHNEIYKIKQALDENKNKFEIK--YQILSDDL 277
QY 302 FLKAYCASAKRVNLEKEFPQSY--VQVVENMIMALISISLDPYNT----- 343
Db 278 I-----SSDKPFLLLFRGEYFNQVNIIDIGIKRVKISRSDINSQPIKQRLLIYYKNV 331
QY 344 -----PMF-----GDSWITDKNFRMAQFASWAR---VFPANQAIK 375
Db 332 YKNFDGNGKEKEVPQFIHEEREISNEYGGHVLNQPFRVAFISSQNNNEVIYVDID 391
QY 376 -----YFATDQKGKAPN-----FLSKALSNAGGYTFTRSCWD--KNA 410
Db 392 NIDNLFDFLEHQQRKKTKITIKTDNTKLTLYEIIYSKELKDSAIKARWENWDPINL 451
QY 411 TVMVL-----KASPPGEFHQAPNGTFELFI-----KGRNFTD-----AGVF 448
Db 452 EHKKLVDKHNGKNGKVPKNPRINPHNGLEKEKIFWLEHEKLNHLSPNFKVFNKEGAF 511
QY 449 V-----YSGD-----EAIMKLRNVQRTRIHSITLTDNQNVITKARQNKW-ETGNL 495

```

```

Db 512 VKVVISQKAYSIDLPPFNSLTLTKYFPMWL---SNMDFQEEELITIKDNNNSFSKSGNFL 568
QY 496 DVLTYTN-----PSYPNLDHQRSVLFINKKYFLVIDRAIGEAATGNLGVHQLK--- 543
Db 569 LELKNTNYSNVLISGVGYNLKLHNL--IDKKILVPVEKSI---AGNILKNYLLKNFN 623
QY 544 -EDSNPVFDKTKNRVYTYTRDGNLMIQSLNADRTSLNEE-----EGKV-----SYVYNKE- 593
Db 624 FBES--ISSLSYBEIVQNNKLLNVLNIEFLNKNKNKIFDIENKMDSLSLTYVDNKNK 681
QY 594 ----LKRPAFVFEKPKKNAGTQNFVSIYVYDQOKAPEISIREKNGDEKGLN---LT 646
Db 682 LLSLDEEFELKINSFN--TKNFVEIDYQD---HPNYDLTFEKSIFYKDKINKKQIT 736
QY 647 LTINGKQQLV 656
Db 737 ININ-KERII 745

RESULT 5
JC5870
poly(beta-D-mannuronate) lyase (EC 4.2.2.3) - Pseudomonas sp.
N:Alternate names: alginate lyase I
C:Species: Pseudomonas sp.
C:Date: 05-Mar-1998 #sequence_revision 13-Mar-1998 #text_change 20-Jun-2000
C:Accession: JC5870
R:Kraiwattapanong, J.; Ooi, T.; Kinoshita, S.
Biosci. Biotechnol. Biochem. 61, 1853-1857, 1997
A:Title: Cloning and sequence analysis of the gene (alyII) coding for an alginate lyase
A:Reference number: JC5870; MUID:98067850; PMID:9404064
A:Accession: JC5870
A:Molecule type: DNA
A:Residues: 1-713 <KRA>
A:Cross-references: DDBJ:AB003330; NID:g2073106; PIDN:BAAL19848.1; PID:g2073107
A:Experimental source: strain OS-ALG-1
C:Comment: This enzyme catalyzes the degradation of alginate by a beta-elimination mecha
cleaved alginate oligomers.
C:Genetics:
A:Gene: alyII
C:Superfamily: Pseudomonas poly(beta-D-mannuronate) lyase
C:Keywords: carbon-oxygen lyase

Query Match 4.4%; Score 154; DB 2; Length 713;
Best Local Similarity 20.3%; Pred. No. 0.0076;
Matches 153; Conservative 92; Mismatches 234; Indels 276; Gaps 42;

QY 87 IRQPIDKVTREMAKALVHQF---QPHKGYGPDYGDINQWMPVKDNEVRWQLHRVKW 143
Db 48 LRQSYQAV-KNAADKALAQPIVVPVFKDGGGYTH-----EQHKNY 88
QY 144 WQAM--ALVYHATGDEKYA-----REWYQYSDWARKNPLGLSQDN-----DKF 185
Db 89 SNMLNCVAYQISGEKKYADYKVNMLNVASQYKWP-LHPKRKSEEDGGRIFWQSLNDF 147
QY 186 VWR--PLEVSDRV-QSLPPTFSLFVNSPAPTEAFLMEF-----LNSYHQADYLSTH 234
Db 148 VMQLYTIQAYDLVYDGIPATDKTIEEKLFPV--ILKFFTEDRDYDFVFNKH----- 196
QY 235 YAEQGNHRLFEAQNLFA-GVS-----FPE--FKDSPRWQRTGISVLNTEIKKQVYA 283
Db 197 -----NHGTW-----NLAAVGITGVNLKREYVEMAIKSGKDKGTGL---AQIDQLFSP 244
QY 284 DGMQFELSPIYHVAIDIFLKAYGSAKRVNLEKEFPQSYVQVVENMIMALISISLDPYNT 343
Db 245 DGYTME-GPYQRYALLPFVL--FAKAIN-NVEPSRKIFEYRDKLSKAIHTSL----- 294
QY 344 PMFGSWITDKNFRMAQFASWARVFPANQAIK----- 375
Db 295 ----QTSYTDKTF-----FPLNDAIKDKTYESVELVYGVLDLAYADIKAEVDLLD 339
QY 376 -----YFATDQKGKAPNLSKAL-----SNAGFYFRSGWKNATV 412

```

Db 340 IARQQNRVIVSDAGLKVRAADLAAGKAVPFYQTLWIRDGKGDEGGILRNGPNTDQC 399
QY 413 MVLKASPPGEFHAQPD-----NGTFELFKRNFPTDAG-----VFVSGDEALMKL 459
Db 400 VVLKASQMGHGHEDRLNLLFYDNTTEIF-----PDYGAARFLNIDTKGGGGLPEN 452
QY 460 RNWYQTRIHTSLTLD-----NQ-NMVTIKARONKWT-----491
Db 453 NTWAKQTVAHNALVVDQTSHEFNALGPADKASPTLLYFSNQPNLKVVSASEDKAYTDVTM 512
QY 492 -----GNLVLVTNPSYNPLDHRQSVLF-----INXKYLVID-----RAIGE 531
Db 513 LRTSALVKEGLDKELLIDVMOQAQSAKSHQYDLPEWYKQQLVNTSPVPTAKANQLTALGD 572
QY 532 ATGNLGVHWLKEDSNPVFDKT-----KNRVYTYRDGN-----LMIOSLNADRTSLNE 581
Db 573 KNGYQHI-W--LNASNPLEGKGMVGLNKNRFYTHFVSDNPLEVKKLUSIGANDPEMNL 629
QY 582 EEGKVSYYVYNKELKRPAPVFEKPKNAGTQNFVSIVPYDG-----OKAPEIS-----629
Db 630 VDGK-----AFMLSSGQN-----QTFVSITEHTGGTDPINETVSSALPTVSGLK 674
QY 630 -IRENKGNDFEKGKLN---LTVTN---GKQQLVLV 658
Db 675 LIKSDAQQTIIISFKVNERTYTYQINYTEKQQLYII 709
RESULT 6
T36084
hypothetical protein SCE134.02c - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T36084
R:Saunders, D.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajadream, M.A.
submitted to the EMBL Data Library, April 1999
Reference number: Z21596
A:Accession: T36084
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-561 <SAU>
A:Cross-references: EMBL:AL049661; PIDN: CAB41200.1; GSPDB: GNO00070; SCOEDB: SCE134.02c
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB: SCE134.02c

Query Match 4.0%; Score 139; DB 2; Length 561;
Best Local Similarity 21.6%; Pred. No. 0.061;
Matches 76; Conservative 45; Mismatches 163; Indels 68; Gaps 12;
QY 273 LNTETIKQVYADGMQFELSPIYHVAIDIFLKAYGSAKRVNLEKEFPQSYQTVTENMIMA 332
Db 186 LAAHMRAAVHRDGEWEASTYVHLFVHLAYLLALRGAR---PERLPDDLAALGAMVSA 241
QY 333 LISISLPDYNTPFGD-----SWITDKNFRMAQFASWAKRVFPAQAIFY 377
Db 242 LAGVAAPDGTLPALHDGYPYRREAOAAETREVAALATQLFAGAPLKSVAQAERDAAALS 301
QY 378 A-TDCKQKQKAPNFLSKA---LSNAGFYFRSGMDKNATVMVLKASPPGEFHAQPDNGTPE 433
Db 302 ADLDWADFDFLPATPAKHTFPDAGIAVFRSA-----GVHALDAGPHGGHGHQKLSLY 357
QY 434 LFI-KGRNFTPDAGVFVYSGDBAIMKLRNWRQTRIHTSLTLTDNQNMVITKARONKWTG 492
Db 358 LVADDGTAWPDGQVPVY---AHRNLRITYASTAAHPTFRVDEREQPCDAVLDD-GDGT 412
QY 493 NNLDLVLTNPSYNPLDHRQSVLFINKKY---FLVIDRAIGATGNLGVHWOLKEDSNPV 549
Db 413 R-----CEQAYDGVGTATRRVV-TDRSYLLDILVLDAA-----443
QY 550 FDKTKNRVTVTYRDGNLMIQSLNADRTSLNEEGKVSYYVYNKELKRPAPVF 601
Db 444 ---TERRLTAQLRPGTDLVVAQGTDRARTVWGDGSAVLTHGHTSRPQAEF 492

RESULT 7
AB1695
p60-related protein homolog lin2100 [imported] - Listeria innocua (strain Clip11262)
C:Species: Listeria innocua
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AB1695
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; M
ok, C.; Schlueter, T.; Simoes, N.; Tietz, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland
A:Title: Comparative genomics of Listeria species
A:Reference number: AB1077; MUID: 21537279; PMID: 11679669
A:Accession: AB1695
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-764 <GLA>
A:Cross-references: GB:AL592022; PIDN: CAC97330.1; PID: g16414614; GSPDB: GNO0178
A:Experimental source: strain Clip11262
C:Genetics:
A:Gene: lin2100

Query Match 4.0%; Score 139; DB 2; Length 764;
Best Local Similarity 19.3%; Pred. No. 0.098;
Matches 148; Conservative 97; Mismatches 280; Indels 242; Gaps 37;
QY 7 KRIL-----VFAVI---ALSSGNILAQSSITRKDFPHINLEYSGLKVKAKVAAAGNYDDAA 60
Db 2 KRILSVLCIFAILVGLFSPFTIKASATNTKTFEEL-----YAEAK 43
QY 61 KAL-LAYREKSKAREPDFSNA-----EKPAIROP-----IDKVTREMADKA 102
Db 44 KHLGKPYSDQDKRRGPNYFDCSGVTQVYKVTGVRIPNTSAPQYSAADKVGKNGKQKPG 103
QY 103 LVHQPQPHKGYGFDYGDINQMWPVKDNEVRWQLHRVKWQMAALVY---HATGDEKTA 160
Db 104 DLVYFKGHVGI-YIGNGRMINAQNDGVKIDNI-----NSSYQSFIVGVGRFNFSEKKG 157
QY 161 REWVTQYSD-----W-----ARKNPLG-----LSQDNDKFWVRPLEVSDRVQSIPPTFS 204
Db 158 SKSAYAVSDNLNRSSNNWDSVAGKVPQAKVSIIDLSDKNGCMVTVNTTKGYMLNTTN 217
QY 205 LFNVSPTAPFAP-----LMEFLNSYHQQADVLSTH 234
Db 218 YFSDTPVITKYAKONINLRKATWDSVAQVKQGEKVTNMLKTNVNGWYQVTVGGKGTG 277
QY 235 YAEQGNHRLFEAQRNL--FAGVSFPEFKDSPRWQTGISVLNTE---IKQVYADGMQPE 289
Db 278 YMLNNYLVENPLNMEYIYAVGTNLRSANW-DSSISLVVPEGRAKVE---MDTN 331
QY 230 LSPYHYA-----AIDIFLKAYGSAKRVNLEKE---FPQSYQTVTE---NMI 330
Db 332 SGPTWKVTYQNTGYIPLTDDYLSKTTVLTKTYAKDNLRKATKATWDSVAQVKQGEKV 391
QY 331 MALISISLPDYNTPMFG-----DSWITDKNFRMAQF-----ASW---AR 366
Db 332 TVNLKTSVNGVQVTVYGGKGMILNDNLYVEKALNMKYIYAVSSILNRSEAKWSSISQ 451
QY 367 VPPANQAIK-----YFAT--DGKQKAP--NFLSKALSNAGFYT-----FRSG 405
Db 452 VVPEGRAVKVEMDITVGNWFKVYDNTKGYMPLNDLYLSETAVLTKYAKONLNRSEAK 511
QY 406 WDKNATVMVLKASPPGEFHAQPDNGTTFELFKRNFPTDAGVFVYSGDBAIMKLRNWRQ 465
Db 512 WDSEVTVQVEKGEKVTVNSKTSIDGWYEV-----TYGKKKGYMILNNY---555
QY 466 TRIHSTLTLDNQNVY--ITKARONKWTG-----NNLDVLTVTNPS 504
Db 556 -LVAEPDLTKYYAVNTLNLRSESKWSSISQVPEGAQKVKVEMNTSDGNWTKYQNTK 614
QY 505 --YPNLDHQRSVLFINKKYFLVIDRAIGATGNLGVHWOLKEDSNPVFDKTKNRVTVTYR 562

Db 615 GYMPNDLNLSETAVLTKYY-----AKDNLNLRSEAKWDS-----EISQVVE 656

QY 563 DGNMLMIQSINADRTSLN-----BEEGKVSYYNKKELKRPFAVPEKP 604

Db 657 KGEKVTINS-----KTSINGWHEVTYGGKGYM-----ILSDNYLVEKP 695

RESULT 8
T44835
hypothetical protein weef [imported] - Acinetobacter lwoffii
C;Species: Acinetobacter lwoffii
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 19-May-2000
C;Accession: T44835
R;Nakar, D.; Gutnick, D.L.
submitted to the EMBL Data Library, July 1999
A;Description: Genomic organization of the wee region of Acinetobacter lwoffii RAG-1 reg
A;Reference number: 222856
A;Accession: T44835
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-604 <NAK>
A;Cross-references: EMBL:AJ243431; PIDN:CAB57203.1
A;Experimental source: strain RAG-1
C;Genetics:
A;Gene: weef
C;Superfamily: Acinetobacter lwoffii hypothetical protein weef

Query Match 3.9%; Score 137.5; DB 2; Length 604;
Best Local Similarity 20.5%; Pred. No. 0.087;
Matches 126; Conservative 76; Mismatches 243; Indels 169; Gaps 33;

QY 90 PIDKVTREMAKALVHQFQPH-----KGYGYFDY-----GKDINQWMP-----128

Db 34 PVOKLSAQLPMGCFQSFQSAFTGMNVFVNSTLTAFGYLKYPVAGLPFNWFSYPLTHIFQ 93

QY 129 -----VKDNEVRQLHRVQWQAMALVYHATGDEKARE-----WV 164

Db 94 NTEKAWYQIPDFGNVDIKGIWEASRFDW--LIDLVLH-----ERQVRQCALVQDLWL 147

QY 165 QYSDWARKNPGLSQDNKDFWR--PLEVSDRVQSLPPTFSLFVNSPFTTP-----AFLM 218

Db 148 ---NDWCERNPAYFGFN-----WKCQGEASIRVWHI---ITALIGLEQEQHPHDNVCAFT 197

QY 219 EFLNSVHQADYLSTHYAEGNHLFEAQNLFAGVSFFPEKDS---PRWRQTGISVLNT 275

Db 198 AHLKRIEPTIDYA---IAQNNHGTSEAAALYIGGAVLNWLKPKSYQKQWALGEXWLLN 254

QY 276 ETKQVYADGMQFELSPIYHVAIDIF-----LKAYGSAKRVNLEKEFPQSYVQ 324

Db 255 RATKLIMQDGGFSQYSINHRVMDLSYCLAETVRQKPFELKPSGQLYRQLKATDMLYVL 314

QY 325 T-----VENM-----IMALLISLPDNTPMFGSDSWITDKNFR--MAQFASNA-----RVFPA 370

Db 315 TQADGVPNLGANDGARLIPVSTQDY-----RDFRPTVQLASTLPHQHSYYPE 362

QY 371 ---NOAIKYFATDGKQKAPNLSKALSACGFTYFRSGW--DKNATVMVLKASPPGEFH 424

Db 363 PDSYDESILTFQI--QKLQDNF---ELPSRQNFNMSGMSQNNFNFIATKLPKFKER 417

QY 425 -AQPNDGTFFELFKGNFTPDAGVFVYSGDEALMKLNWYQRIHSTLTLDNQNMV--I 481

Db 418 PSQCDALHLDVFKGNLRLDGGTYSN---STVEDLNPFGVSASHTVEFQHQMPRL 474

QY 482 TYARQNKWTGNMLDLVLTNPSYP-----NLDHORSVLFINKKYFLVIDRAIGATG 534

Db 475 SRFLFGAWLTPKE---LITYSANEFQGYQDHWGCTHHRKI-----SLTDNAIKADN 523

QY 535 NLG-----VHMOLKED-----SNPYFDKTKNRVYTYRGNLIMTOSLNADR--TSINEE 583

Db 524 ISGFQQQAVLRWLQPDHWTLENNLSNGKVERII-----LEAHPITLRLWSE 570

QY 584 GKVSYYNKKELKRP 597

Db 571 GBESRYYYQSKVP 584

RESULT 9
E96554
hypothetical protein F19C24.18 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 15-Jun-2001
C;Accession: E96554
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Pederspiel, N.A.; Kaul, S.; White, O.; Alonso, C.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: E96554
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-560 <STO>
A;Cross-references: GB:AB005173; NID:gl1094759; PIDN:AAG29692.1; GSPDB:GN00141
C;Genetics:
A;Gene: F19C24.18
A;Map position: 1
C;Superfamily: human mannosyl-oligosaccharide 1,2-alpha-mannosidase

Query Match 3.6%; Score 127; DB 2; Length 560;
Best Local Similarity 20.2%; Pred. No. 0.43;
Matches 89; Conservative 82; Mismatches 159; Indels 110; Gaps 22;

QY 7 KRIIVFAVIALSSCNILAQSSITRK---DFDHINLEYGLEKVKVAKVAGNVDDRAKAL 63

Db 26 RRLALLFIVFVSVMVLDNRINLAREHEVEFVKLENEVRLEQMLBELNGGVGNKPLKTL 85

QY 64 LAYREKSKAREPDFSNAEKPADIQPIDKVTREMAKALVHQFQPHKGYGYFDYGKDN 123

Db 86 -----KDAPED---PVDKQRQKVKEMIAHWSSYEKA---WQKD-- 120

QY 124 WQWMP--VKDNEVRQLHRVQWQAMALVYHATGDEKY--AREWVVOYSDWARKNPLGLSQ 180

Db 121 -ELQPRTKDGTDSFGGLGATMVDLSLTLYIMGLDEQFKAREWVASSLDFDKDYDASME 179

QY 181 -----DNDK--FYWRPLEVSDRVQSLPPTFSLFVNSPFTPAFLMEFLNSY 224

Db 180 TTIRVVGGLLSAYDLSGDKNFKLEKAKDIADR---LLPAW---NTPTGIPYNIILRNG-- 231

QY 225 HQADYLSTH---YAEQGNHRLFEAQNLFAGVSFFPEKDSPRWRQTGISVLNTEIKQV 281

Db 232 -----NAHNFSWAAGGDSILADSSTEQLFIALSORTGDPKYQKQKVKVI--TELKNRP 283

QY 282 YADGMQFELSPIY-----HVAIDIF-----LKAYGSAKRVNLEKEFPQSYV 323

Db 284 PADG---LLPTIYNPDNANPSYSTTTFCAGMGSFVEYLLKVVQGNKTSVAKPYRDWME 339

QY 324 OTVENMIVALISLIPDYNTPMFGSDSWITDKN-----FRMAQFASWARVFPANQAIKYP 378

Db 340 KSMKG--LUSLVKKSTPSSFT-----YTCEKNGNLLDKMDLACFA---PGMLAL----- 385

QY 379 TDGKQKAPNLSKALSNAK 398

Db 386 --GASGYGDEBKFLSLAG 403

RESULT 10
AC0422
hypothetical protein YP03474 [imported] - Yersinia pestis (strain CO92)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001

C:Accession: AC0422
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
Hill, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AC0422
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-672 <KUR>
A:CROSS-references: GB:AL590842; PIDN:CAC92703.1; PID:gi5981398; GSPDB:GN00175
C:Genetics:
A:Gene: YPO3474

Query Match 3.6%; Score 125.5; DB 2; Length 672;
Best Local Similarity 18.9%; Pred. No. 0.73;
Matches 114; Conservative 81; Mismatches 248; Indels 159; Gaps 24;

QY	87	IRQPIDKVTREMAKALVHQ	-----FOPHKG-----YGYF-DYKXDNQMW-----PV 129
DB	17	WQOP-EIATLADNQVLETLVPATGIATWNHYICPEHGVLWDRFSPQAHRCV 75	
QY	130	KDNEVRWOLHRVKNQA	-----MALVYHATGDEKY---AREWVYQSDWARKNP 175
DB	76	DDHLFSGEPYDGAWWRALNGLNAKACNLGLLQTLTGEIRYDKVRDILMSY---ARYYP 132	
QY	176	LGLSDNDKFWVRPLEVSDRVOSLPPTSLFVNSAFIPA	-----FLMEPLNSYHQQA 228
DB	133	-----DYEVHGIP-----YNGFGKANAQTLCEANCLLDPALGYDFTA 170	
QY	229	DYLSHYAEQGNHRLFEAQRNLFAGVSF	-----PEFKDSPRWQRTGISV----- 272
DB	171	DTLSQ-----EQRCLERLRVGADFLMQHRTQLHNEVKISAIIVIGLIEEHY 224	
QY	273	-----LNTIEIKQVYADGMQFELSPYHYAAIDIFLKAYGSAKRVNLEKFPQSY 322	
DB	225	IEFAVNADYGLAYQLEHGLFNEGLWFEQSVHYHYA---LQGFWSFEKLAAGSRYSILA 280	
QY	323	VOTVENMIMALLISLDPYNTFMFGDSWTDKRNFRMAQFASWARVFPAN	----- 371
DB	281	LPYRDMLSFPLKLLMPNGTFPRINDCTAGQQLNAHLYEFAYQIYQVREYAAALQHIY 340	
QY	372	-----QAIKYPATDGKQKAPNFLSKALSNAGFYTFRSGMDKNAT 411	
DB	341	ROPRINLDALLYGAEPLQIIDVPTDLH--AP-----DCGLTILQ--POASR 388	
QY	412	VNVLKASPPGEPAQPDNGTFFELFKGRNFTPDAGVYVYSGDEAIMKLRNWRQTRIHT 471	
DB	389	ALLIKHSPYGGHHDYRLNLILFDRGHEVLPDLGTTGYA---QRHYDYKNSATHNT 444	
QY	472	LFLDNQNVITKARQNKWETGN---LDVLTNPSYPNLDHORSVLFINKKXFLVIDRA 528	
DB	445	LSINQKNPPAVPYTRQWHQANFSLDTEVDWROTPAPELDSHTRVQWDEASY-----RD 499	
QY	529	IGBATGNLGVHQLKEDSNPVDKTKNRVYTYRDCGNLMIOSLNADRTSLNEEGKVS 588	
DB	500	VKERRILLWLEDTIIDISINVPHQSQWNTLYIDGT---ATEQKGMTSFG-DSGPMQY 555	
QY	589	VY 590	
DB	556	VH 557	

RESULT 11
G82875
hypotheical protein U0558 [imported] - Ureaplasma urealyticum
C:Species: Ureaplasma urealyticum
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: G82875
R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.
submitted to GenBank, February 2000
A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a mid

A:Reference number: A82870
A:Accession: G82875
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1883 <GLA>
A:CROSS-references: GB:AB002154; GB:AF222894; NID:G6899557; PIDN:AAF30971.1; GSPDB:GN00175
A:Experimental source: serovar 3; biovar 1
C:Genetics:
A:Gene: U0558
A:Genetic code: SGC3

Query Match 3.6%; Score 124.5; DB 2; Length 1883;
Best Local Similarity 20.1%; Pred. No. 4.3;
Matches 118; Conservative 74; Mismatches 200; Indels 195; Gaps 31;

QY	192	VSDRVQSLPPTFSLFVNSPAPTPAFLEFLNSYHQQADYLSHYAEQGNHRLFEAQRNL 251	
DB	631	VTDRY-----PYLSLVAQS---DPAFKTTGIN-----YLKIVSTH---EYGHQTLQDMKDIS 677	
QY	252	-----AGVSFPPEF-----KDSPRWQRTGISVLNTEIKKQVYADGMQFELS 291	
DB	678	DSNDSVIGGGIDSRSGVDSYNGKALQDYLNAESSGLTFKTDVYNPTKDGSEFNFS 737	
QY	292	-----PIYHVAADIFLKAYGSAKRVNLEKFPQSYVQTVENMIMA----- 332	
DB	738	LNNDPKNPVWETQK-DIF-----GSVNADDPKFAFFYNNKRRFLQKDELFEAAKLNRVQPY 792	
QY	333	-LISISLDPYNT---PMEGDSWTDKRNFRMAQFASWARVFPANQAIKYFATD--GKQK 385	
DB	793	DLFINMSFDHESATVNPSPGPD-INKPSRLKAEY-----YFYNDQNSQEQK 837	
QY	386	APNF-LSKALSNAGFYTF---RSG-----WDKNATVMVLKASPPGEPAQPDNGTFFEL 434	
DB	838	NDNFKFGSVIEKPGSLKYDGILKDCMGTPIKFSKQGRALVYLKHKHKKPKYKDD---IBI 894	
QY	435	FTKGRNFTP---DAGVYVYS-GDEAIMKLRNWRQTRIHTLTLDNQNVITKARQN-KWE 490	
DB	895	LIKTNTPTVIDLSTCLSKSDGTINTRKLNKVRBIQ-----DSINSILVKNYNGGWD 947	
QY	491	TGNLDV-----LVTNPSYPNLDHORSVLFINKK----- 520	
DB	948	ENGNDFTSMFNLTVDHPMTSNKRWAEIRITAMFKYDPFNLSKSLKDEKANASSV 1007	
QY	521	--YFLVIDRAIGEAATGNLGVHQLKEDSNPVDKTKNRVYTYRDCGNLMIOSLNADRTS 578	
DB	1008	PYKVKLSLIQS-----DIINKTDFQKYALLTLQYNASKIWDNKSKEVEVK 1055	
QY	579	LINEEGKVSIVYNKELKRPAP-----VPEKPKQKQAGTON 612	
DB	1056	LDPELIEIKKYDYDKQFDRFGVKNNVVAQTTLTFNYFDAGIEGNNGYKYFVKPKKELYQN 1115	
QY	613	FYSIVYPYDQKAPK-----ISIR-----ENKGNDFE---KGKLN 645	
DB	1116	IIR-----TKTESVESIIIGIRFSYLTQNKITSYEQLFNKSMLN 1155	

RESULT 12
IPEY
beta-fructofuranosidase (EC 3.2.1.26) 2 precursor - yeast (Saccharomyces cerevisiae)
N:Alternate names: beta-D-fructofuranoside fructohydrolase; invertase; protein YII162w
C:Species: Saccharomyces cerevisiae
C:Date: 13-Jun-1983 #sequence_revision 13-Jun-1983 #text_change 21-Jul-2000
C:Accession: A00899; S50365; S24557; S05871; S05870; S59658; A27748; A45656
R:Tausig, R.; Carlson, M.
Nucleic Acids Res. 11, 1943-1954, 1983
A:Title: Nucleotide sequence of the yeast SUC2 gene for invertase.
A:Reference number: A00899; MUID:83168934; PMID:6300785
A:Accession: A00899
A:Molecule type: DNA
A:Residues: 1-532 <TAU>
A:CROSS-references: EMBL:V01311; NID:G3833; PIDN:CAA24618.1; PID:G3834
R:Lyne, G.; Bowman, S.; Churcher, C.
submitted to the EMBL Data Library, December 1994

[illegible]

QY 387 -----PNFLSKALSNAGFYTFPSGWDK---NATV---MVLKASPPGEBFAQPDNGTTEL 434
DB 226 QNKSNNPFGNYASDIKYANSQDTRGLNSAIDEIPLFKIDP--RTHSHFKNWFD- 282
QY 435 FIKGRNFTPDAGVYVSGDEALMKLRNRYOTRIHSTLTLDNQNMVITKARQKWETGNN 494
DB 283 ---RNFTKEKIQYYSQELVPR-----EFQKIAPVYNSQKT----- 318
QY 495 LDVLYTTPSYNLDHQRSVLFIKKYFLVIDRATGEATGN-LGVHWQJKE--DSNPVFD 551
DB 319 -----TNPELNEDDE-----FSQDGFGLVRLKLTSSVTPNAPAFDWSRTALINGPTLH 367
QY 552 KTKN-----RVVTVYRDG---NNLMIQSLNADRTSLNEE 582
DB 368 KNENVVLGDNFFLNPPYGGNIYVFKDIYGVSPRTYNEKAFMKNVTWNEFYNAFLEAMN-Y 426
QY 583 EGVKSIVY-----NKLKRPAPVFEKPKKXNAGTQNFVSVIYPYDGGQKAPESIRENKGND 637
DB 427 KSEISQLFVKNESNLKSKAKV--NPGKNNHVRFTGFLTKKAKKAKYKALVLTDEKGN 484
QY 638 FEKGLKNTLTITNG 651
DB 485 KRQSNLKF-LTRNG 497

RESULT 14

E70533
probable sulfatase (EC 3.1.6.-) atSB - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: E70533
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skeiton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: E70533
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-970 <COL>
A:Cross-references: GB:296070; GB:AL123456; NID:G3261791; PIDN:CAB09444.1; PID:e320915;
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: atSB
C:Keywords: sulfuric ester hydrolase
F:258/Modified site: 3-oxoalanine (Cys) #status predicted

Query Match 3.4%; Score 120.5; DB 2; Length 970;
Best Local Similarity 20.0%; Pred. No. 2.9;
Matches 85; Conservative 60; Mismatches 150; Indels 129; Gaps 22;
QY 131 DNEVRQLHRVKKWQAMA--LVYHATGDEKYAREWYQVSDWARKNPLGLSQNDKPVWR 188
DB 386 DKALEW-LHTVRAQNAKTPWMLYATG-ATHAPHVFK--EWADKYRGEPDGDGDVYRQK 441
QY 189 PLEVSRRVQLPTFSLFVNSPAPTFAPLMEFLNSYHQADYLSHYAEQGNHRLFEAOR 248
DB 442 TFERQRLGIIPDPAEL-TERPDLPFAW-----DSMS-----EAQRLLFARQM 483
QY 249 NLFAGVS-----FPFQKDSPR-----WRQTGISVLNT---EIKQVYADGM 286
DB 484 EVFAGSEADNWNVGRLLDALEDLGESDNTLVFYINGDASMEGINTGSENFMTFLNGL 543
QY 287 -----QFELSPIY-HVAAI-----DIFLKAYGSAKRVNLE--KEPPQSYVQVTENMIM 331
DB 544 DLDAERQLELIEQYGGIAALGDEFTAPHFASAHASNTPLQWKGQWASHLGGTRDPLV 603
QY 332 A-----LISISLDP-----YNTPMFGDSWITDKNPRMA 359
DB 604 ANPARIKPDGRVRSQFTHCIDIAPTVLAIGLPHTHVDGFEQEPMDGTSEFV----- 655

QY 360 QFASWARVFPANQA-----IKFYATDGGKGKAPNLFSLKALSNAAGFYTF-----RSGWDKWA 410
DB 656 -----RTFDDAARHRTVQYF-----ENFGSRAIYKDGWACARLQKAPWDLSP 700
QY 411 TMMVLKASPPGEBFAQPDNGTTELFIKGRNFTPDAGVYVSGDEALMKLRNRYOTRIHS 470
DB 701 ETM--RRFAPGTY--DPQDVMELYLPDDFSAQKLAELAEHPDKVAELTQLWQBAERNR 756
QY 471 TLTL 474
DB 757 VLPL 760

RESULT 15

C90593
hypothetical protein MYPV_6510 [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C:Species: Mycoplasma pulmonis
C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C:Accession: C90593
R:Chambaud, I.; Hellig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.
Nucleic Acids Res. 29, 2145-2153, 2001
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pu-
A:Reference number: A9512; MUID:21267165; PMID:11353084
A:Accession: C90593
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1272 <KUR>
A:Cross-references: GB:AL445566; PID:G14090066; PIDN:CAC13824.1; GSPDB:GN00153
A:Experimental source: strain UAB CTIP
C:Genetics:
A:Gene: MYPV_6510
A:Genetic code: SGC3

Query Match 3.4%; Score 120; DB 2; Length 1272;
Best Local Similarity 20.1%; Pred. No. 4.9;
Matches 139; Conservative 96; Mismatches 218; Indels 238; Gaps 40;

QY 77 DFSN-AEKPADIRQPIDKVTREMAKALVHQFQPHKGYGYDYDGK-DINQWMPVKDNEV 134
DB 558 DFSNKKDLASSVWMDKIIKE---SLVVK---NKTENFDNFKYDISVSSSL-DEV 608
QY 135 RQOLHRVKKWQAMALVYHATGDEKYAREWYQVSDWARKNPLGLSQNDKFWRPLEVSD 194
DB 609 NGKL-KIK-----MTIFKTKDR--LKEFLYEVVGF---QLGITKN----- 644
QY 195 RVQSLPTFSLFVNSPAPTFAPLMEFLNSYHQADYLSHYAEQGNHRLFEAQ---RNL 250
DB 645 -----LISYRNDKLSADQIIREFYVLGNATPEQLEKLSL 683
QY 251 FAGVSPE-----FKDSPRWRTGI-----SVLNTETKKQVYADGMQF 288
DB 684 FS-VAIPDGLPEPEFISFKAKANSNDRGILTYTLRKNKTNARSGKINVEKTHNIIDSYLA 742
QY 289 ELSPIYHVAI-DIFLKAYGSAKRVNLEKEPPQSYVQVTENMIM-----ALIS----- 335
DB 743 Q-----HVLVNEIALENDSPFKNSINELSPN---QILENITLFDKNKSIITPEKDIOV 794
QY 336 ---ISLPDYNTPM--FGDSWITDKNFRMAQF-----ASWARVFPANQAIKYFA 378
DB 795 KYKIANPQYNLQNSINVEIFEKGHSQKVTRELVNGPKFKFDASWFDVVAKNLIKDNP 854
QY 379 TDGKQKQKAP-NFLSKALSNAAGFYTFRSQWMDKNAITMVL-----KASPPGEFHA 425
DB 855 RDIKQDKTHFKFIANTFGTKSFY-FK--FQNNQTSKELLKHVNSVDNIEYKNVAGEIH- 910
QY 426 QPDNGTTEL-FIKGRNFTPDAGV---VYSGDEALMKLRNRY---QTRHSTLTLDN- 476
DB 911 -----FDVKFTKSSSSSETNIVIRKVFSGFKDLIFDTEQYDDEDEHRSIATTK 964
QY 477 --QNMVITK---ARQNKWETGNNLDVLT--YTNPSPFNLDHORSVLFIKKYFLVIDRAI 529
DB 965 MPRNQFIQKIVEARQN-----NDLDALVDVFTNLIAPFRISNETL-----SFSIRQNM 1013

```
QY 530 GEATGNLGVHWOLKED-----SNPVEDKTKN----- 555
Db 1014 GSPRKNKGAH---REDIADQTAEQFTILHRLSNLVTNKEKNIRDHIREWLNETVFAIRTS 1070
QY 556 RVYTTYRDGNNLMIQSLNADRTSLNEE-----EGKSVYVYNKELKRPAPVFEKPKKNA 608
Db 1071 HLTSEFPKNEFLAKTSSERIELIKKYVDLYIPEGYTLNIYN----- 1113
QY 609 GTQNFVSVIYVPYDGGOKAPEISIRENKGNDPE 639
Db 1114 -THNITSLKYKYNQVE-----KENKLSDFE 1138
```

Search completed: July 28, 2003, 20:53:09
Job time : 58 secs

This Page Blank (uspto)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 4, 2003, 10:23:59 ; Search time 393 Seconds
(without alignments)
3459.332 Million cell updates/sec

Title: US-09-802-285A-2

Perfect score: 3494
Sequence: 1 MTIKFKRIIVFAVIALSG.....KGKMLTLTINGKQQLVLVP 659

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 1439767 seqs, 1031500376 residues

Total number of hits satisfying chosen parameters: 2879534

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DRV=xlh
-Q=/cgn2_1/USPTO.spool/US09802285/runat_28072003_190514_9775/app_query.fasta_1.839
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09802285@cgn_1_1_221_@runat_28072003_190514_9775
-NCPU=6 -TCPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:*

1:	/cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2:	/cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3:	/cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4:	/cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5:	/cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6:	/cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7:	/cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8:	/cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9:	/cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10:	/cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq1:*
11:	/cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
12:	/cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq3:*
13:	/cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
14:	/cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
15:	/cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
16:	/cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
17:	/cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	166.5	4.8	11427	11	US-09-070-927A-165 Sequence 165, App

C	2	116	3.3	4702	8	US-08-781-986A-268	Sequence 268, App
	3	116	3.3	14121	12	US-09-802-640-31	Sequence 31, Appl
	4	116	3.3	14121	12	US-09-920-033-3	Sequence 3, Appl
	5	115.5	3.3	3996	11	US-09-801-368-369	Sequence 369, App
	6	115	3.3	3222	11	US-09-974-300-2654	Sequence 2654, App
	7	115	3.3	1830121	15	US-10-329-960-1	Sequence 1, Appl
	8	114.5	3.3	1830121	15	US-10-329-960-1	Sequence 1, Appl
	9	113.5	3.2	4765	15	US-10-090-624-5	Sequence 5, Appl
	10	112.5	3.2	6720	11	US-09-070-927A-321	Sequence 321, App
	11	111	3.2	2253	11	US-09-993-292A-19	Sequence 19, Appl
	12	111	3.2	8908	11	US-09-993-292A-18	Sequence 18, Appl
	13	110.5	3.2	1689	10	US-09-815-242-4610	Sequence 4610, Ap
	14	110.5	3.2	1737	10	US-09-815-242-8108	Sequence 8108, Ap
	15	110.5	3.2	10194	11	US-09-070-927A-91	Sequence 91, Appl
	16	110	3.1	14070	11	US-09-870-759-127	Sequence 127, App
	17	110	3.1	14635	11	US-09-880-107-2287	Sequence 2287, Ap
	18	109	3.1	32480	11	US-09-847-101B-23	Sequence 23, Appl
	19	109	3.1	34427	12	US-09-111-911-5	Sequence 5, Appl
	20	109	3.1	35408	15	US-10-155-649-3	Sequence 3, Appl
	21	109	3.1	35871	10	US-09-956-335-2	Sequence 2, Appl
	22	109	3.1	35935	10	US-09-725-720-43	Sequence 43, Appl
	23	109	3.1	35935	11	US-09-782-378A-4	Sequence 4, Appl
	24	109	3.1	35935	11	US-09-782-378A-5	Sequence 5, Appl
	25	109	3.1	35978	10	US-09-956-335-1	Sequence 1, Appl
	26	109	3.1	36620	12	US-09-952-060-30	Sequence 30, Appl
	27	109	3.1	37474	12	US-09-952-060-25	Sequence 25, Appl
	28	109	3.1	38519	12	US-09-952-060-28	Sequence 28, Appl
	29	108	3.1	4938	15	US-10-055-794-1	Sequence 1, Appl
	30	108	3.1	6525	15	US-10-055-794-3	Sequence 3, Appl
	31	107.5	3.1	4011	12	US-09-934-455-21	Sequence 21, Appl
	32	107.5	3.1	4011	15	US-10-278-173-127	Sequence 127, App
	33	107	3.1	5926	11	US-09-070-927A-220	Sequence 220, App
	34	106	3.0	1703	11	US-09-974-300-644	Sequence 644, App
	35	106	3.0	4896	10	US-09-740-274-3	Sequence 3, Appl
	36	106	3.0	10317	10	US-09-849-866-1	Sequence 1, Appl
	37	105.5	3.0	3504	10	US-09-815-242-7425	Sequence 7425, Ap
	38	105.5	3.0	19031	11	US-09-070-927A-167	Sequence 167, App
	39	105	3.0	2979	15	US-09-842-484A-1	Sequence 1, Appl
	40	105	3.0	2979	15	US-10-184-485-4	Sequence 4, Appl
	41	105	3.0	3540	12	US-09-759-130B-41	Sequence 41, Appl
	42	104	3.0	5547	15	US-10-242-056-48	Sequence 48, Appl
	43	104	3.0	7551	10	US-09-817-514A-1	Sequence 1, Appl
	44	104	3.0	7551	15	US-10-242-056-46	Sequence 46, Appl
	45	103.5	3.0	3827	11	US-09-870-759-44	Sequence 44, Appl

ALIGNMENTS

RESULT 1

US-09-070-927A-165/c
; Sequence 165, Application US/09070927A

; Patent No. US20020120116A1

; GENERAL INFORMATION:

; APPLICANT: Charles A. Kursch

; Patrick J. Dillon

; Steven Barash

; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides

; NUMBER OF SEQUENCES: 982

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/070,927A

; FILING DATE: 04-May-2000

CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 60/046,655
 FILING DATE: 1997-05-16
 APPLICATION NUMBER: 60/044,031
 FILING DATE: 1997-05-06
 APPLICATION NUMBER: 60/066,009
 FILING DATE: 1997-11-14
 ATTORNEY/AGENT INFORMATION:
 NAME: Kenley K. Hoover
 REGISTRATION NUMBER: 40,302
 REFERENCE/DOCKET NUMBER: PB369
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 309-8504
 TELEFAX: (301) 309-8512
 INFORMATION FOR SEQ ID NO: 165:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 11427 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 SEQUENCE DESCRIPTION: SEQ ID NO: 165:
 US-09-070-927A-165

Alignment Scores:
 Pred. No.: 3,31e-07 Length: 11427
 Score: 166.50 Matches: 121
 Percent Similarity: 35.18% Conservatives: 89
 Best Local Similarity: 20.27% Mismatches: 215
 Query Match: 4.77% Indels: 174
 DB: 11 Gaps: 26

US-09-802-285A-2 (1-659) x US-09-070-927A-165 (1-11427)

QY 120 LysAspIleAsnTyrPheValAsnGluValArgTyrPheLeuHis 139
 Db 7871 AAAGATGCTGCGAATCGGTATCTGATGATGATCCAGATGGCTCTTATGTAGT 7812
 QY 140 ArgValLysTyrPheGlnAlaMetAlaLeuValTyrHisAlaThrGlyAspGluLysTyr 159
 Db 7811 CGTCAAGCTTCTTGTAGATCGGCACAGCATATGCATTTACTTAAAGAAAGCGTTAC 7752
 QY 160 AlaArgGluTyrValTyrGlnTyrSerAspTyrAlaArgLysAsnProLeuGlyLeuSer 179
 Db 7751 TTACAGAAATGGCAGCGTCTGCTTATGATTTATT-----AACGATGAGGTGAGCCA 7698
 QY 180 GlnAspAsnAspLysPheValTyrArgProLeuGluValSerAspArgValGlnSerLeu 199
 Db 7697 AATTCGACGAATAGGATGTTGGCTCGTTAGATGTTGGATTCGAGTAACAACAAAC--- 7641
 QY 200 ProProThrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGlu 219
 Db 7640 -----TGATGAAAAGCTTGACGTATATTTCCAAATCGCTGATTTTCA 7599
 QY 220 PheLeu-----AsnSerTyrHisGlnGlnAlaAspTyrLeuSer 232
 Db 7598 CTATTAGGAATGTAGTGTGTGAACAACGCCCTTGTGATCCATCTGGATTTTGGAG 7539
 QY 233 ThrHisTyrAlaGluGlnGlyAsnHisArgLeuPheGlu----- 245
 Db 7538 CGGTCTTATATCGATAA-----TACAGGTCTAGTAATTTGGGTGTCTTGGCAATTTGGT 7485
 QY 246 -----AlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPheLysAspSerPro 263
 Db 7484 GGAATGGCAGCTATTGATTTATT-----CTTCAGAACTGGTGACCACT--- 7440
 QY 264 ArgTyrArgGlnThrGlyLysValLeuAsnThrGluLysLysGlnValTyrAla 283
 Db 7439 AAACAGAGGATCTAATATGCTCTGTTACTGAACAGCTGTGATCTACAAATTTCTATCA 7380
 QY 284 AspGlyMetGlnPheGluLeuSerProIleTyrHisValAlaAlaIleAspIlePheLeu 303
 Db 7379 GATGGAAATTCATTGGGAGCAGACCCCGCTGTAC----- 7347

QY 304 LysAlaTyrGlySerAlaLysArgValAsnLeuGluLysGluPheProGlnSerTyrVal 323
 Db 7346 -----CAGCAGAAAGTTTGTGATGACATTCGTG 7320
 QY 324 GlnThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuPro----- 339
 Db 7319 TATCTATTGCAG-----ATTCTGAATATCTTGAAGTCAGTACCATTAGATCTTCGC 7266
 QY 340 ---AspTyrAsnThrProMetPheGlyAspSerTyrPheThrAspLys----- 354
 Db 7265 ATGAAATTAACAAACACCTATTTCTCTACCCATTATTTGGCGATAACCAAGATATCTTA 7206
 QY 355 -----AsnPhe----- 356
 Db 7205 AATCCGATCAATGATAGTATGATCATGTCACTTTTCATTACGTATATATATATATCGCAAA 7146
 QY 357 -----ArgMetAlaGlnPheAlaSerTyrPheAla 365
 Db 7145 TTAGGTTTCATATTTGAACCTCCATGACTGCGAATATGGCAAGCTT-----TGGACG 7092
 QY 366 -----ArgValPheProAlaAsnGlnAlaIleLysTyrPheAla--- 378
 Db 7091 GGGGATCTTTATGAAGAAAGATCTGGAAAC-AATGAAGCCAAAGAACTTTTCGTGG 7033
 QY 379 -----ThrAspGlyLysGlnGlyLysAlaProAsnPheLeuSerLysAlaLeu 394
 Db 7032 CGAATCAAGTGGGCTGATGGCTACAAAGCAGA-----GGATATCTTA--- 6991
 QY 395 SerAsnAlaGlyPheTyrThrPhe---ArgSerGlyTyrAspLysAsnAlaThrValMet 413
 Db 6990 -----TTTTACATTTTAAACGGTCTGCATGG----- 6964
 QY 414 ValLeuLysAlaSerProProGlyGlu-PheHisAlaGlnProAspAsnGlyThrPheGlu 433
 Db 6963 -----GAGCGCATGCTCATGCTTCTACAGGTGGATTTAC 6928
 QY 433 uLeuPheIleLysGlyArgAsnPheThrProAspAlaGlyValPheValTyrSerGlyAs 453
 Db 6927 ACTACACTACAAAGGATGACTTATTTCCGATAGTGTGCTGTACAGCTATGTCAACAA 6868
 QY 453 pGluAlaIleMetLysLeuArgAsnTyrTyrArgGlnThrArgIleHisSerThrLeuTh 473
 Db 6867 ATCAGAGCGACTTCAGCTA-----AAAGAGTGGCTTCGCACAATACCATGTT 6820
 QY 473 rLeu---AspAsnGlnAsnMetValIleThr-----LysAlaArgGlu 486
 Db 6819 TATCGCAAAAATCCCATACCTTTAGTTTCCGATACGTGGGGTTATGACAAATTACCAC 6760
 QY 486 nAsnLysTyrPheGluThrGlyAsnLeuAspValLeuThrTyrThrAsnProSerTyrPr 506
 Db 6759 ACCCTTATTTCCAGCAATAAAGAACTTTCTGTGCTGTTTTTTCGAGAATCGGCTGGCT 6700
 QY 506 oAsnLeuAspHisGlnArgSerValLeuPheIleAsnLysLysTyrPheLeu----- 523
 Db 6699 GGATAAGGCGGATCAGAATCCAATGATTTTGGCGGAGCTTCATCTATTTAAAGTCGAT 6640
 QY 524 -----ValIleAspArgAlaIleGlyGlu-----AlaThr 533
 Db 6639 CAATCCGTAGTTATTATTGATAGCTTTGACAGGACAGAAAGAGACTGAAATTTACAGTAC 6580
 QY 533 rGlyAsnLeuGlyValHisTyrPheGlnLeuLysGluAspSerAsnProValPheAspLysTh 553
 Db 6579 CTATAATTTGGCACCCTCGATAAATTTGTCAAAAGAGAGCGCATCGGTTTGTCTTAACATC 6520
 QY 553 rLysAsnArgValTyrThrTyrArgAspGlyAsnAsnLeuMetIleGlnSerLeuAs 573
 Db 6519 GAACAAGCATTAAGTACACACTCTCTTTTGTCTGAGGCG----- 6483
 QY 573 nAlaAspArgThrSerLeuAsnGluGluGlyLysValSerTyrValTyrAsnLysGlu 593
 Db 6482 -----CAACACAGCAATCAGTGGCAAAAGGTTTCAGAGATTTTAAACCACT 6436

QY 593 uLeuLysArgPro-----AlaPheValPheGluLysProLysLysAsnAlaG1 609
DB 6435 GAATGAGCACCACCGCTTATCAAAAGTTTGTACAAACCGG-AAAGAAATTCAG 6377
QY 609 yThrGlnAsnPheValSerIleValTyrProTyrAspGlyGlnLysAla 625
DB 6376 CAACAGTCATTCTCGGTAGAGATACCAATACGCCAATAAAGT 6328

RESULT 2
US-08-781-986A-268/c
; Sequence 268, Application US/08781986A
; Publication No. US20030054436A1
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 268:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4702 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-781-986A-268

Alignment Scores:
Pred. No.: 0.0428 Length: 4702
Score: 116.00 Matches: 138
Percent Similarity: 33.20% Conservative: 103
Best Local Similarity: 19.01% Mismatches: 301
Query Match: 3.32% Indels: 185
DB: 8 Gaps: 27

US-09-802-285A-2 (1-659) x US-08-781-986A-268 (1-4702)

QY 4 LysIlePheLysArgIleIleValPheAlaValIleAlaLeuSerSerGlyAsnIleLeu 23
DB 3000 AAATATTCAGGAGGTTCTAGTTATGGCAATGATTAAAGTAGTCCAGGGAATCAGA 2941
QY 24 AlaGlnSerSerIleThrArgLysAspPheAspHisIleAsnLeuGluTyrSerGly 43
DB 2940 GCAAAATCGCAATCTTACGGGCAAGGTCA--GACCAATCCGTCAAATTTTATCTGAT 2884
QY 44 LeuGluLysValAsnLysAlaValAlaAlaGlyAsnTyrAspAlaAlaLysAlaLeu 63
DB 2883 TTAACMCGTGCACAAGGTGAATTCGACG--AACTGGGAAGTCAAGCT----- 2836
QY 64 LeuAlaTyrTyrArgGluLysSerLysAlaArgGluProAspPheSerAsnAlaGluLys 83

DB 2835 TTCAGCCGTTTCGAGAGCAATTCACAACTAGTCTCT-----AAAGTAGAAAA 2785
QY 84 ProAlaAspIleArgGlnProIleAspLysValThrArgGluMetAlaAspLysAlaLeu 103
DB 2784 TTTGCACAATATTAGAAAGAAATTAACAACAATGAATAGCACTGCTGATCCGTTCAA 2725
QY 104 ValHisGlnPheGlnProHisLysGlyTyrGlyTyrPheAspTyrGlyLysAspIleAsn 123
DB 2724 GAACAAGACCAACAACACTTCTTAATAATTTCGGT----- 2692
QY 124 TrpGlnMetTrpProValLysAspAsnGluValArgTrpGlnLeuHisArgValLysTrp 143
DB 2691 -----TTGCAATAAGCATTCGAATTTGCCAAAGTCACAT-TTCTTAATG 2648
QY 144 TrpGlnAlaMetAlaLeuValTyrHisAlaThrGlyAspGlu-----LysTyrAlaArg 161
DB 2647 TGGCTTTGCTTATCATTTTTTAAGAAAAACAACACTGAAAGGAAATAAGCATGAAAAAGAA 2588
QY 162 GluTrpValTyr----- 165
DB 2587 AATTCGATTTATGCATTAATTCCTTAATTATTATTATTAATGCGCATGTTAGTAGATA 2528
QY 166 -----GlnTyrSerAspTrpAlaArgLysAsnProLeuGlyLeuSerGln 180
DB 2527 TTTTGTTCACAAACAATATCGAGATCAATCAGAAAGAGGATCCCAAGTGTAGTAAT 2468
QY 181 AspAsnAspLysPheValTrpArgProLeuGluValSerAspArgValGlnSerLeuPro 200
DB 2467 AAAAAATAATAA-----ATACATATCGCAATTTGTTAACGAGGATCAACCA 2423
QY 201 ProThrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGluPhe 220
DB 2422 ACGACATATAAC-----CGTAAAAAGTTGAGCTGGGTCAAGCATTTATTAAAGTTA 2369
QY 221 LeuAsnSerTyrHisGlnGlnAlaAspTyrLeuSerThrHisTyrAlaGlu----- 237
DB 2368 GCAATAGAAAAACTATAAATTTGAAACAGTAAACAAGAACGTTGCTGCTGCTGTTG 2309
QY 238 ---GlnGlyAsnHisArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyValSerPhe 256
DB 2308 AAAAAATGGTGGATACCAAGTCATGTTGTTATCCCAAGAAAACTTTTCAAAATGGCAATG 2249
QY 257 ProGluPheLysAspSerProArg-----TrpArgGlnThrGlyIleSer--- 271
DB 2248 CAATTAGACGCTAAACACCACTCGAAATATCGTACAGTATATAAACACCTGTAGACAA 2189
QY 272 -----ValLeuAsnThrGluIleLysLysGlnValTyrAlaAspGlyMetGln 287
DB 2188 AAAGAAGAGTAGCTAAAAACACACAGAAAAAGTTGTAAGTAATGCTACTTAACGACTTAA 2129
QY 288 PheGluLeuSerProIleTyrHisValAlaAlaIleAsp----- 300
DB 2128 AAAAACTTAGTCGAATTTTATTAAACAGCATCATTTGATAATTTACATATATGCACAAAA 2069
QY 301 -----IlePheLeuLysAlaTyrGly-----SerAlaLysArg 311
DB 2068 AATGTGGCGCTATTATGACGCGTGAAACATGGTGGAATAGTAATTCGAAATTA 2009
QY 312 ValAsnLeuGluLysGluPheProGlnSerTyrValGlnThrValGluAsnMetIleMet 331
DB 2008 TTAATTCCAATTAAACGACTTCCCGGAATTTATTACAGATACGCTGTAAATTCATTTCT 1949
QY 332 AlaLeuIleSerIleSerLeuProAspTyrAsnThrProMetPheGlyAspSerTrpIle 351
DB 1948 GCAAAACAAAGACATTACA----- 1931
QY 352 ThrAspLysAsnPheArgMetAlaGlnPheAlaSerTrpAlaArgValPheProAlaAsn 371
DB 1931 ----- 1931
QY 372 GlnAlaIleLysTyrPheAlaThrAspGlyLysGlnGlyLysAlaProAsnPheLeuSer 391
DB 1931 ----- 1931

```
Db 1930 -----AAATGTTCCAAACA-----TACAAAT 1910
QY 392 LysAlaLeuSerAsnAlaGlyPheTyrThrPheArgSerGlyTyrAspLysAsnAlaThr 411
Db 1909 AAATCATTTAGTGGCGAATTCAGATAGCTTCAGAGTCAAGTCAACAGACAGATTATATGTTTCG 1850
QY 412 ValMetValLeuLysAlaSerProProGlyGluPheHisAlaGlnProAspAsnGlyThr 431
Db 1849 ACTTTAAATTGAAAA-----CAAAATTCATTA 1823
QY 432 PheGluLeuPheIleLysGlyArgAsnPheThrProAspAlaGlyValPheValTyrSer 451
Db 1822 TTITGAC-----GAGCAAAATACAGCGATCGATGATAAAATGTTACAGATTATATAA 1775
QY 452 GlyAspGluAlaIleMetLysLeuArgAsnTrpTyrArgGlnThrArgIleHisSerThr 471
Db 1774 TCGCAAAAAGATAGCGTGAACCTTGATACTAT-----ATCAATGCATTA 1730
QY 472 LeuThrLeuAspAsnGln-----AsnMetValIleThrLysAlaArgGln 486
Db 1729 AAACAGATGGACAGCAAAATTGATCAACAATCAAGTATGCAAGATACAGGTAAAGAGAA 1670
QY 487 AsnLysTrpGluThrGlyAsnAsnLeuAspValLeu-----ThrTyrThr 501
Db 1669 TATTAACAACACTGTTAAAGAAAACCTTAGATAAATTAAGAGAAATCATTCATCAACAGAG 1610
QY 502 AsnPro-----SerTyrProAsnLeuAspHisGlnArgSerValLeuPheIleAsnLysLys 520
Db 1609 TCACCATTTTCAAAGGTATGATGCAAGANTATCGTAAGCAATTAACAGATCATCTGCA 1550
QY 521 TyrPheLeuValIleAspArgAlaIleGlyGluAlaThrGlyAsnLeu----- 536
Db 1549 GATGAGCTTGCAATTAACAAGACTTACAAGATGCGCTAAATAGCATTAAAAATGAACAAT 1490
QY 537 -----GlyValHisTrpGlnLeuLysGluAspSerAsnProValPheAsp 551
Db 1489 GCTCAATTCGCTGAAAACCTTAGAGAAACAACCTTCATGATGATATGTCACAAAGAACCTGAT 1430
QY 552 LysThrLysAsnArgValTyrThrTyrArgAspGlyAsnAsnLeuMetIleGlnSer 571
Db 1429 ACAGATACAACATTTATCTAT-----AACATGCTTAACAACAGAC 1391
QY 572 LeuAsnAlaAspArgThrSerLeuAsnGluGluGluGlyLysValSerTyrValTyrAsn 591
Db 1390 TTTATAGCT-----GCAGGTGTTAAATGAGGTGAAGCTAATAATAACAGCAATGTC 1337
QY 592 LysGluLeuLysArgProAlaPheValPheGluLysProLysLysAsnAlaGlyThrGln 611
Db 1336 AAAGACCAAAACGTTATAAAATGAATATTAATTTGAAAAACCGTTAGCAGAACACATT 1277
QY 612 AsnPheValSerIleValTyrProTyrAspGlyGlnLysAlaProGluIleSerIleArg 631
Db 1276 AATTTAAACAGAT-----TACGATAACCAAGTTGCGCAAGACACAAGTAGTTTG 1229
QY 632 GluAsnLysGlyAsnAspPheGluLysGlyLysLeuAsnLeuThrLeuThrIleAsnGly 651
Db 1228 ATTAATGATGTGTCAAAGTGCACAGCTACTGAA-----ACGATTAAAAAGTAATGAT 1178
QY 652 LysGlnGlnLeuValLeu 657
Db 1177 ATTAATCAATTAACCTGTT 1160
```

RESULT 3

```
US-09-802-640-31
; Sequence 31, Application US/09802640
; Publication No. US20030036057A1
; GENERAL INFORMATION:
; APPLICANT: Braun, Andreas
; APPLICANT: Bonsal Aruna
; APPLICANT: Kleyn Patrick
; TITLE OF INVENTION: GENES AND POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISEASE AND THEIR USE
; FILE REFERENCE: 24736-2048
```

```
; CURRENT APPLICATION NUMBER: US/09/802,640
; CURRENT FILING DATE: 2001-03-09
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: FastSeq For Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 14121
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (129)...(13820)
; OTHER INFORMATION: Nucleotide sequence encoding apolipoprotein B
; OTHER INFORMATION: (APOB)
US-09-802-640-31
```

Alignment Scores:

Pred. No.:	0.236	Length:	14121
Score:	116.00	Matches:	104
Percent Similarity:	30.9%	Conservative:	94
Best Local Similarity:	16.28%	Mismatches:	197
Query Match:	3.32%	Indels:	244
DB:	12	Gaps:	26

US-09-802-285A-2 (1-659) x US-09-802-640-31 (1-14121)

```
QY 218 MetGluPheLeuAsnSerTyr-----HisGlnGlnAlaAspTyr 230
Db 9330 ATAGACTTCTCGAATACTATGCACTGTTCTGAGTCCAGCGCCACGAAAGTGTGG 9389
QY 231 -----LeuSerThrHisTyrAlaGluGlnGlyAsn 240
Db 9390 CAAGTAAGTGTAGTTCATCAGTATAAGTACACCAAAATTTCTCTGCTGGAACAC 9449
QY 241 HisArgLeuPheGluAla-----GlnArgAsnLeu----- 250
Db 9450 GAGAACATTATGAGGCCCTATAGGAATAAATGAGAGCAAAATCTGGAATTTCTTAAAC 9509
QY 251 -----PheAlaGlyValSerPheProGluPhe 259
Db 9510 ATTCTTTAACAANTCTGAAATCGTCTACCTTACACAATAATCACAACTCCTCCACTG 9569
QY 260 LysAspSerProArgTyrArgGlnThrGly----- 269
Db 9570 AAAGATTCTCTCTATGCGAAAAAAGAGCTTTGAAGGAATTTCTTGAACACGCAAGCAA 9629
QY 269 ----- 269
Db 9630 TCATTGATTTAAGTGTAAAGCTCAGTATAAGAAAAACAAACACAGCAGCATTCATCACA 9689
QY 270 -----IleSerValLeuAsnThrGluIleLysLysGlnValTyrAlaAspGlyMetGln 287
Db 9690 AATCCTTTGGCTGTGCTTTGTGAGTTTATCAGTCAGACATCAAAATCCTTTGACAGGAT 9749
QY 288 PheGluLeuSerProIleTyrHisValAlaAlaIleAspIlePheLeuLysAlaTyrGly 307
Db 9750 TTTCAAAAAAAC-----AGAAACAATGCATTAGATTTTGTCAACCAATCTATAAT 9800
QY 308 SerAlaLys-----ArgValAsnLeuGluLys-----GluPheProGln 320
Db 9801 GAAACAAAAAATTAAGTTTGTAGTACAAAGCTGAAAAATCTCAGCAGCAGCTCCCGAGG 9860
QY 321 SerTyrValGlnThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuProAsp 340
Db 9861 ACCTTT----- 9878
QY 341 TyrAsnThrProMetPheGlyAspSerTyrIleThrAspLysAsnPheArgMetAlaGln 360
Db 9879 TACACTGTTCCAGTTGTC-----AATGTTGAAGTGTCTCCA 9914
QY 361 Phe-----AlaSerTrpAlaArgValPheProAlaAsnGlnAlaIleLysTyr 376
Db 9915 TTACCATAGATGTCGGCATTCGGTATGTGTCTCCAAAGACAGTCAGCATCCTAGT 9974
```

```
QY 377 PheAlaThrAspGlyLysGlnGlyLysAlaProAsnPhe----- 389
Db 9975 TTCTCCATCCTAGGTTCTGAGCGCGTGTCTTATACACATTAATATCGGCATCATTA 10034
QY 389 ----- 389
Db 10035 GAGTCGCCAGTCCTTCATGTCCTAGAAATCTCAAGCTTCTCTTCCACATTTCAAGGA 10094
QY 390 -----LeuSerLysAlaLeuSerAsnAlaGlyPhe-----TyrThr 401
Db 10095 TTGTGTACCATAGCCATATTTTATTCCTGCCATGGCAATATTACCTATGATTTCTCC 10154
QY 402 PheArgSerGlyTrpAspLysAsnAlaThrValMetValLeuLysAlaSerProGly 421
Db 10155 TTTAAATCAAGT-----GTCAACACTGAAATACCAATGCT 10190
QY 422 GluPheHisAlaGlnProAspAsnGlyThrPheGluLeuPheIleLysGlyArgAsnPhe 441
Db 10191 GAACITTTTAAACCAAGTCAGAT-----ATTGTTGCTCATCTCTTTCATCTTCATCT 10244
QY 442 ThrProAspAlaGlyValPheValTyrSerGlyAspGluAlaIle----- 456
Db 10245 GTCAATTGATGCACTGCAAGTAAATAGAGGGCACCACCAAGATTGCAAGAAAAAGGGGA 10304
QY 457 -----MetLysLeuArgAsnTyrArgGlnThrArgGlnIleHisSer 470
Db 10305 TTGAAGTTAGCCACAGCTCTGCTCTGAGCAACAAATTTGTGAGGGGTAGTCATAACAGT 10364
QY 471 ThrLeuThrLeuAspAsnGlnMetValIleThrLysAlaArgGlnAsnLysTyrGlu 490
Db 10365 ACTGTGAGCTTAAACCGAATAATATGGAAGTGTCAAGTGCAGCAAAACACCAAAAGCCGAA 10424
QY 491 -----ThrGlyAsnAsnLeuAspValLeu 498
Db 10425 ATTCCAATTTTGAGATGAATTTCAAGCAAGAACTTAATGGAATACCAAGTCAAAACCT 10484
QY 499 ThrTyrThrAsnProSerTyrProAsnLeuAspHisGlnArgSerValLeuPheIleAsn 518
Db 10485 ACTGTCTCTTCCATGGAATTTAGTATGATTTCAATTTCTCAATGCTGTACTTACC 10544
QY 519 LysLys-----TyrPheLeuLeu 524
Db 10545 GCTAAAGGAGCAGTTGACCACAAAGCTTAGCTTGGAAGCCCTCACCTTCTTCTTCCATT 10604
QY 525 IleAspArgAlaIleGlyGluAlaThrGlyAsnLeu----- 536
Db 10605 GAGTCATCTACCAAGAGAGATGTCAAGGGTTCGGTTCTTCTCGGGAATATTCAAGAACT 10664
QY 536 ----- 536
Db 10665 ATTGTAGTGAGGCCAACACTTACTTGAATTCACAGAGCACAGGTCTTCAGTGAAGCTG 10724
QY 537 ---GlyValHis-----TyrPheLeuLysGluAspSerAsnProValPhe 550
Db 10725 CAGGGCACTTCCAAATTTGATGATATCTGGAACCTTCAAGTAAAGAAATTTTGTCTGGA 10784
QY 551 AspLysThrLysAsnArgValTyrThrTyrArgAspGly-----AsnAsnLeuMet 568
Db 10785 GAAGCCCACTCCACAGCATATATTCCTCTGGAGCAGCAGTACGAAAAACCACTTACAG 10844
QY 569 IleGlnSerLeu-----AsnAlaAspArgThrSer-----LeuAsnGluGlu 582
Db 10845 CTAGAGGGCTCTTTTTCACCAACGGAGACATACAGCAAGACCCCTGGAACTCTCT 10904
QY 583 GluGlyLysValSerTyrValTyrAsnLysGluLeuLysArgProAlaPheValPheGlu 602
Db 10905 CCATGGCAATGTGTCAGCTCTTGTTCAGGTCCATCAAGTCAGCCAGTCTCTTCCATGAT 10964
QY 603 LysPro-----LysLysAsnAlaGlyThrGlnAsn----- 612
Db 10965 TTCCTGACCTTGCCCGAGGAGTGGCCCTGAATGCTAACATGATAGAACCAAGAGATCAGA 11024
QY 613 PheValSerIleValTyrProTyrAspGlyGlnLysAlaProGluIleSerIleArgGlu 632
```

```
Db 11025 TGAATAAATGAAGTCGGATTCTTCTGGTCTTTCCAGGCCAGGTCGAGCTT----- 11078
QY 633 AsnLysGlyAsnAspPheGluLysGlyLysLeuAsnLeuThrLeuThrIleAsnGly 651
Db 11079 -----TCCAATGACCAAGAAAAGGCACACCTTGACATTGCAGGATCTCTTAGAAGGA 11129

RESULT 4
US-09-920-033-3
; Sequence 3, Application US/09920033
; Publication No. US20030087853A1
; GENERAL INFORMATION:
; APPLICANT: Rosanne M. Crooke
; APPLICANT: Mark J. Graham
; TITLE OF INVENTION: ANTISENSE MODULATION OF APOLIPOPROTEIN B EXPRESSION
; FILE REFERENCE: ISPH-0592
; CURRENT APPLICATION NUMBER: US/09/920,033
; CURRENT FILING DATE: 2001-08-01
; NUMBER OF SEQ ID NOS: 123
; SEQ ID NO 3
; LENGTH: 14121
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (129)...(13820)
US-09-920-033-3

Alignment Scores:
Pred. No.: 0.236 Length: 14121
Score: 116.00 Matches: 104
Percent Similarity: 30.99% Conservative: 94
Best Local Similarity: 16.28% Mismatches: 197
Query Match: 3.32% Indels: 244
DB: 12 Gaps: 26

US-09-802-285A-2 (1-659) x US-09-920-033-3 (1-14121)

QY 218 MetGluPheLeuAsnSerTyr-----HisGlnGlnAlaAspTyr 230
Db 9330 ATAGACTTCTCGAATAACTATGCACTGTTTCTGAGTCCCAAGTCCCGCAGCAAGCAAGTTGG 9389
QY 231 -----LeuSerThrHisTyrAlaGluGlnGlyAsn 240
Db 9390 CAAGTAAGTCGTAGTTCAATCAGTATAGTACACCAAAATTTCTCTGCTGGAACAAC 9449
QY 241 HisArgLeuPheGluAla-----GlnArgAsnLeu----- 250
Db 9450 GAGCAACATTATGAGGCCCATAGTAAGTAATAATGAGAGCAAAATCTGGATTCTTTAAAC 9509
QY 251 -----PheAlaGlyValSerPheProGluPhe 259
Db 9510 ATTCTTTTAACAATTCCTGAAATGCGTCTACCTTACACATAATATCACAATCTCTCCACTG 9569
QY 260 LysAspSerProArgTyrArgGlnThrGly----- 269
Db 9570 AAAGATTTCTCTCTATGCGGAAAAAACAAGCTTGAAGGAATCTTGAAGAACGACAAGCAA 9629
QY 269 ----- 269
Db 9630 TCATTGTTAATTAAGTGTAAAGCTCAGTATAGAAAAAACAACACAGGCATTCATCACA 9689
QY 270 -----IleSerValLeuAsnThrGluIleLysGlnValTyrAlaAspGlyMetGln 287
Db 9690 AATCTCTTGTGCTGCTGCTTTGTGAGTTTATATCAGTCAGAGCATCAAAATCTTTGACAGCAT 9749
QY 288 PheGluLeuSerProIleTyrHisValAlaAlaIleAspIlePheLeuLysAlaTyrGly 307
Db 9750 TTTGAAAAAAC-----AGAAACATGCAATTAGATTTTGTCCACCAATCTATAAT 9800
QY 308 SerAlaLys-----ArgValAsnLeuGluLys-----GluPheProGln 320
Db 9801 GAAACAAAAATTAAGTTTGATAAGTACAAAGCTGAAAAATCTCACGACGAGCTCCCCAGG 9860
```

```
QY 321 SerTyrValGlnThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuProAsp 340
DB 9861 ACCTTT-----AsnAlaAspArgThrSer-----LeuAsnGluGlu 582
QY 341 TyrAsnThrProMetPheGlyAspSerTrpIleThrAspLysAsnPheArgMetAlaGln 360
DB 9879 TACACTGTTCCAGTTGTC-----AATGTTGAAGTGTCTCCA 9914
QY 361 Phe-----AlaSerTrpAlaArgValPheProAlaAsnGlnAlaIleTyr 376
DB 9915 TTCACCATAGAGATGTCGGCAITTCGGCTATGTTCTCCAAAAGCAGTCAGCATGCTAGT 9974
QY 377 PheAlaThrAspGlyLysGlnGlyLysAlaProAsnPhe----- 389
DB 9975 TTCTCCATCTAGTTCTGACGTCGCTGTCATACACATTATTCCTGCCATCATTA 10034
QY 389 ----- 389
DB 10035 GAGTGCCAGTCCTTCATGTCCTAGAAATCTCAAGCTTTCTCTCCACATTTCAAGGAA 10094
QY 390 -----LeuSerLysAlaLeuSerAsnAlaGlyPhe---TyrThr 401
DB 10095 TTGTGTACCATAGCCATATTTTATCTGCAITGGGCAATATACCTATGATTTCTCC 10154
QY 402 PheArgSerGlyTrpAspLysAsnAlaThrValMetValLeuLysAlaSerProProGly 421
DB 10155 TTTAAATCAAGT-----GTCAATCACACTGAATACCAATGCT 10190
QY 422 GluPheHisAlaGlnProAspAsnGlyThrPheGluLeuPheIleLysGlyArgAsnPhe 441
DB 10191 GAACTTTTAAACCATGTCAGT-----ATTGTTGTCTCATCTCTCTTCTCATCTTCATCT 10244
QY 442 ThrProAspAlaGlyValPheValTyrSerGlyAspGluAlaIle----- 456
DB 10245 GTCATGTATGACTCGACTGACATAAATAGAGGCGCACCAAGATTGACAGAAAAAGGGA 10304
QY 457 -----MetLysLeuArgAsnTrpTyrArgGlnThrArgIleHisSer 470
DB 10305 TTGAAGTTAGCCACAGCTCTGTCTCTGAGCAAAATTTGTGGAGGTAGTCATAACAGT 10364
QY 471 ThrLeuThrLeuAspAsnGlnAsnMetValIleThrLysAlaArgGlnAsnLysTrpGlu 490
DB 10365 ACTGTGAGCTTAACACGAAAAATATGAAGTGTCAAGTGGCAAAACCAACCAAGCCGAA 10424
QY 491 -----ThrGlyAsnAsnLeuAspValLeu 498
DB 10425 ATTCCAATTTTCAGATGAATTTCAAGCAAGACTTAATGGAATATACCAAGTCAAAACCT 10484
QY 499 ThrTyrThrAsnProSerTyrProAsnLeuAspHisGlnArgSerValLeuPheIleAsn 518
DB 10485 ACTGTCTCTCTCCATCGAATTTAAGTATGATTTCAATTTCTCAATGCTGTACTCTACC 10544
QY 519 LysLys-----TyrPheLeuVal 524
DB 10545 GCTAAAGAGAGAGTTGACACCAAGCTTAGCTTGGAAAGCCTCACCTCTTACTTTTCCAT 10604
QY 525 IleAspArgAlaIleGlyGluAlaThrGlyAsnLeu----- 536
DB 10605 GAGTCATCTACAAAGGAGATGTCAGAGGTTTCGGTCTTTCTCGGGAATATTCAGGAAC 10664
QY 536 ----- 536
DB 10665 ATTGCTAGTGGGCCAACACTTACTTGAATTTCCAGAGCACACGGTCTTTCAGTGAAGCTG 10724
QY 537 ---GlyValHis-----TrpGlnLeuLysGluAspSerAsnProValPhe 550
DB 10725 CAGGCACTTCCAAATATGATGATATCTGGAACCTTGAAGTAAAGAAAAATTTTCTCGGA 10784
QY 551 AspLysThrLysAsnArgValTyrThrTyrArgAspGly-----AsnAsnLeuMet 568
DB 10785 GAAGCCACACTCCAAACGCATATATTCCCTCTGGGAGCACAGTACGAAAAACCACTTACAG 10844
```

```
QY 569 IleGlnSerLeu-----AsnAlaAspArgThrSer-----LeuAsnGluGlu 582
DB 10845 CTAGAGGGCTCTTTTTCACCAACGGAGAACTATACAGCAACCCCTGGAACTCTCT 10904
QY 583 GluGlyLysValSerTyrValTyrAsnLysGluLeuLysArgProAlaPheValPheGlu 602
DB 10905 CCATGCGAAATGTGAGCTCTTGTTCAGGTCCATGCAAGTCAGCCAGTTCCTTCCATGAT 10964
QY 603 LysPro-----LysLysAsnAlaGlyThrGlnAsn----- 612
DB 10965 TTCCCTGACCTTGGCCAGGAAGTGGCCCTGTAATGCTAACTAAGAACAGAGATCAGA 11024
QY 613 PheValSerIleValTyrProTyrAspGlyGlnLysAlaProGluLysSerIleArgGlu 632
DB 11025 TGGAAATGAAGTCCGATTCATTTCTGGTCTTTCCAGAGCCAGTCCAGCTT----- 11078
QY 633 AsnLysGlyAsnAspPheGluLysGlyLysLeuAsnLeuThrLeuThrIleAsnGly 651
DB 11079 -----TCCATGATCAACAGAAAGGCACACCTTGACATTCAGAGATCCTTAGAAGGA 11129
```

RESULT 5

US-09-801-368-369
; Sequence 369, Application US/09801368
; Patent No. US20020128250A1

GENERAL INFORMATION:

; APPLICANT: Busby, Robert
; APPLICANT: Call, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 369

; LENGTH: 3996
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-369

Alignment Scores:

Pred. No.: 0.0378 Length: 3996
Score: 115.50 Matches: 157
Percent Similarity: 31.21% Conservative: 108
Best Local Similarity: 18.49% Mismatches: 287
Query Match: 3.31% Indels: 297
DB: 11 Gaps: 45

US-09-802-285A-2 (1-659) x US-09-801-368-369 (1-3996)

```
QY 1 MetThrThrLysIlePheLysArgIleValPheAlaValIleAlaLeuSerSerGly 20
DB 73 ATCTCAAGTAAGCTCGTGAAGAGATTATCAAGTCATTCAGCCATAAATCAAGATCT 132
QY 21 AsnIleLeuAlaGlnSer-----SerSerIleThrArg 31
DB 133 GATTGAAAGCATTTGGGTGGTTCGGAACAATAAGACGCGCCAGTCAGTGTACTTTT 192
QY 32 LysAspPheAspHisIleAsnLeuGluTyrSerGlyLeuGluLysValAsnLysAlaVal 51
```

193 AAGGAC---CGATACGTTTTCATGATCGCTATATCTGAAAGCTAAAAAGACCGCT 249
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 QY 52 AlaAlaGlyAsnTyrAspAspAlaAlaLysAlaLeuLeuAlaTyrTyrArgGluLysSer 71
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 250 TTAGATGACTACTACACGAGGGGCATAAACTCACTAACCGCTACGAGGAAGACGCGT 309
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 QY 72 LysAlaArgGluProAspPheSerAsnAlaGluLysProAlaAspIleArgGlnProIle 91
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 310 GATGACGAATATTTCGGTTGCTAATGGC----- 339
 QY 92 AspLysValThrArgGluMetAlaAspLysAlaLeuValHisGlnPheGlnProHisLys 111
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 340 GACAGAATTGATGACGACCTG-----CACTCA 366
 QY 112 GlyTyrGlyTyrPheAsp-----TyrGlyLysAspIleAsnTrpGlnMetTrpPro 128
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 367 GGTGTCGAAGTTTTCCTCCACTACACCTTATTGCGAGAAATGAGG----- 411
 QY 129 ValLysAspAsnGluValArgTrpGln-----LeuHisArgValLysTrpTrpGln 145
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 412 TCAGACAGTATGCTAGCTTGGATGAAATGCGACCGAATGCGTTCAAATGGCAGTCA 471
 QY 146 AlaMetAlaLeuValTyrHis-----AlaThrGlyAspGluLysTyrAlaArgGlu 162
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 472 ATGCTGGCCAGAGTGTGTAAGGGAGATATTGTTAAAGGTGAAGACGAGGATTGCTAAC 531
 QY 163 TrpValTyrGlnTyrSerAspTrpAlaArgLysAsnProLeuGlyLeuSerGlnAspAsn 182
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 532 CRAAGTC-----AAGAAACCA---GGGTAAATAAGGAGCTC 564
 QY 183 AspLysPheValTrpArgProLeuGlu-----ValSerAspArg 195
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 565 TCAGATGAGTATGCTCGAATGAGGCATGCTGAATGGGAGGACCATGCAAGAGATG 624
 QY 196 ValGlnSerLeu-----ProProThrPheSerLeuPheValAsnSerProAla 211
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 625 GAACAGTCGCTTACATATTAAAGAGATAGTTCAGATTCGTTTGAAGAGATAATGAAG 684
 QY 212 Phe-----ThrProAlaPheLeuMetGlu 219
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 685 TTTCAAATCCACAGGGCAAGATATTGAGCTGGATGCACTGGAGGCGCATCTACCAAGAC 744
 QY 220 PheLeuAsnSerTyrHisGlnAlaAspTyr-----LeuSerThrHisTyrAla 236
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 745 CTCATGAACAGATATCACCGGTGCTCTTATGGCTTAACITGAAGAAAAATGTATAAG 804
 QY 237 GluGln-----GlyAsnHisArgLeuPheGluAlaGln----- 247
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 805 GATAAACCAATACCAATACTGCAGAAATTTACCGCTAGATAAGACGTAATGAATCTTGG 864
 QY 248 -----ArgAsnLeu 250
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 865 CTGAACCTTTAAACGAACTTAACGTTTGAGGAGCAAGAGTTGGACGACTGGATAAACCGT 924
 QY 251 PheAlaGlyValSerPhe-----ProGluPheLysAspSerProArgTrp 265
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 925 TTCTCACCATAAGTAGTTCGGATAATGCCAGAGAGATTTGATGGTGGCCCAATGG 984
 QY 266 ArgGlnThrGlyLysSerValLeuAsnThrGluIleLysGlnValTyrAlaAspGly 285
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 985 ---AAGTGCAGAAATGAAGATCTTCGAGAACAAATGATGAAGGAAGAAACATCGAGTCT 1041
 QY 286 MetGlnPheGluLeuSerProIleTyrHisValAlaAlaIleAspIlePheLeuLysAla 305
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 1042 ATA---TTCCAAAAAATTTTCTATCCGCTATCATCCTTGGATGTTCAAACTGAAACTA 1098
 QY 306 TyrGlySerAlaLysArgValAsnLeuGluLys----- 316
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 1099 CATTTTATAGTCTACAGAGAACTTTGACAAAGATGAACATAAATATCCTTTATGAAGG 1158
 QY 317 -----GluPhePro-----Gln 320
 ||||| : : : : : ||||| : : : : : ||||| : : : : :

1159 TTAAGATCACTACTGGGTTCCCGCTCTATTATTAATCAAGAAGCTATTTTGCATGATTG 1218
 QY 321 SerTyrValGlnThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuProAsp 340
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 1219 TCATATGCACGAAGCTTAAATCCAAATGATGATGATGATGATGATGATGATGATGATGAT 1278
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 QY 341 TyrAsnThr----- 343
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 1279 TTTAACGCTTTTANTTCGACTTTCTGTGCAATTGAAGTACACACTGACAAAATATGCTCC 1338
 QY 344 -----ProMetPheGlyAspSerTrpIleThrAspLys 354
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 1339 AATTTCGCGTTCGATGCGATTTCGACCGGCTTCGAAATACTGTAATA----- 1389
 QY 355 AsnPheArgMetAlaGlnPheAlaSerTrpAlaArgValPheProAlaIleGlnAlaIle 374
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 1390 -----GAAGCCATT 1398
 QY 375 LysTyr-----PheAlaThrAspGlyLysGlnGly---Lys 385
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 1399 CGTTATTTTATTTTCTGTGTAATTTAAAGTTGATTCAGTTCAGTAAACAAAATTTCAAA 1458
 QY 386 AlaProAsnPheLeuSerLys-----AlaLeuSerAsnAlaGlyPheTyrThrPhe 402
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 1459 GCACCCGATCTACTCTTGAATACTGGGATCACCTAAACAAACACCGGTCACTATATT-- 1515
 QY 403 ArgSerGlyTrpAspLysAlaThrValMetValLeuLysAlaSerProProGlyGlu 422
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 1516 -----AACGGTCAGAAACCGTGATT-----CCAAATGAA 1545
 QY 423 PheHisAlaGlnProAspAsnGlyThrPheGluLeu-----PheIleLysGlyArg 439
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 1546 TTCTCAAGTTTAACCTTTGAGACTCGTACATAATTCGAATTCATCTTTTGAAACAAACA 1605
 QY 440 AsnPheThrProAspAlaGlyValPheValTyrSerGlyAspGluAlaIleMetLysLeu 459
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 1606 AACTTCCCAACCAATTTGCT-----AACGCTCAGAAGCA----- 1641
 QY 460 ArgAsnTrpTyrArgGlnThrArgIleHisSerThrLeuThrLeuAspAsnGlnAsnMet 479
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 1642 GAAAAATGG-----CTAAGTTCCATTTTCGAAAAATTTG 1674
 QY 480 ValIleThrLysAlaArgGlnAsnLysTrpGluThrGlyAsnAsnLeuAspValLeuThr 499
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 1675 GGTGCCATGAAAGAAAGCTGAACAGTTT-----AGCAATATCTTAGTCAAGGCG 1725
 QY 500 TyrThrAsnProSerTyrProAsnLeuAspHisGlnArgSerValLeuPheIleAsnLys 519
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 1726 TTCCAAAATTCGCTGTTTATCAGATTATCATATGCAACACTTGT-----AAA 1776
 QY 520 Lys-----TyrPheLeuValIleAspArgAlaIleGlyGluAlaThrGly 534
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 1777 AAGTAAAAAGATGCTCCTATTTTGGTATATCTCC-----GGTAACACTTTTGAG 1827
 QY 535 AsnLeuGlyValHisTrp-----GlnLeuLysGluAspSerAsnProVal 549
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 1828 TCTAGTGGTGTAATATGTTTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1887
 QY 550 PheAspLysThrLysAsnArg-----ValTyrThrThrTyrArgAspGly 564
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 1888 TTAAGAATTTTCGAAATAAATCCATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1947
 QY 565 AsnAsnLeuMetIleGlnSerLeuAsnAlaAspArgThrSerLeuAsn----- 580
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 1948 AATAATTTCAATGTGTATGATATACACAAAGAAACAGATTTGAACATCTTAGTATCG 2007
 QY 581 -----GluGluGluGlyLysValSerTyr-----ValTyrAsnLysGluLeuLys 595
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 2008 AAGGGGAGGATTTCCAAAGGAATTCCTTACTACCGAGTAGTAGCAAAATTCGTCAAGTGAT 2067
 QY 596 ArgProAlaPheValPheGluLysProLysLysAsnAlaGlyThrGlnAsnPhe----- 613
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 2068 TTGGCAGCGATGCTCATCAGTCCAAAAGAGAAATTTTTCACAGACCTTTTGATCAG 2127
 ||||| : : : : : ||||| : : : : : ||||| : : : : :

QY 613 ----- 613

Db 2128 CACCTTGATGAAGAACAAATGAGTTTGAATGGAAGTTCCTTTGAGCTCATTTGGGT 2187

QY 614 ---ValSerIleValTyPro-----TyrAspGlyGlnLysAlaProGlu 627

Db 2188 GCACCTAGTTGTACTATATCTCGAGAGCCAGTAGTTTGGGATGGA-----CCAGTA 2238

QY 628 IleSerIleArgGluAsnLys-----GlyAsnAspPheGluLysGlyLysLeu--- 643

Db 2239 TATAAGCTTCCAGGTAACAACCTTTTTCATCCAAACGAATGATTTAGGGAAAAATGGT 2298

QY 644 -----AsnLeuThrLeuThrIleAsn 650

Db 2299 AACCCAAATACCTTGATTTTACTCAAT 2325

RESULT 6

US-09-974-300-2654

; Sequence 2654, Application US/09974300

; Patent No. US20020146721A1

; GENERAL INFORMATION:

; APPLICANT: Berka, Randy M.

; APPLICANT: Clausen, Ib Groth

; TITLE OF INVENTION: Methods For Monitoring Multiple Gene

; TITLE OF INVENTION: Expression

; FILE REFERENCE: 10085.500-US

; CURRENT APPLICATION NUMBER: US/09/974,300

; CURRENT FILING DATE: 2001-10-05

; PRIOR APPLICATION NUMBER: 09/680,598

; PRIOR FILING DATE: 2000-10-06

; PRIOR APPLICATION NUMBER: 60/279,526

; PRIOR FILING DATE: 2001-03-27

; NUMBER OF SEQ ID NOS: 8481

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2654

; LENGTH: 3222

; TYPE: DNA

; ORGANISM: Bacillus licheniformis

US-09-974-300-2654

Alignment Scores:

Pred. No.: 0.0308 Length: 3222

Score: 115.00 Matches: 147

Percent Similarity: 30.93% Conservative: 85

Best Local Similarity: 19.60% Mismatches: 252

Query Match: 3.29% Indels: 267

Db: 11 Gaps: 36

US-09-802-285A-2 (1-659) x US-09-974-300-2654 (1-3222)

QY 15 IleAlaLeuSerSerGlyAsnIleLeuAlaGln---SerSerSerIleThrArgLysAsp 33

Db 1360 ATAGCAGAGATACAGGGGAATATGTTGGCTCAATGGAGACGGCTCTTACTCGAAGCAAG 1419

QY 34 PheAspHisIleAsnLeuGluTyrSerGlyLeuGluLysValAsnLysAlaValAlaAla 53

Db 1420 GACGAGCAAAAT-----AAAATTTCAAGGAAGCTGAAGAA 1455

QY 54 GlyAsnTyrAspAspAlaAlaLysAlaLeuLeuAlaTyrTyrArgGluLysSerLysAla 73

Db 1456 GAATCCTCTAATCTATCTGCTAAACAGGCTGCTGTTGTGAAAAAATAGTAATAAGCA 1515

QY 74 ArgGluProAspPheSerAsnAlaGluLysProAlaAsp----- 86

Db 1516 AAAGAAAAACAAATAAAGCGCCGAAGAAGACGGCTGATTCGTTATTGCTGCTGCCGAT 1575

QY 87 -----IleArgLysProIleAspLys-----Val 94

Db 1576 GATCAGTATTATGATAAAGGCACAAATTTCAAGAAAAAGACGACGATGACACAGTTGGAAAA 1635

QY 95 ThrArgGluMetAlaAspLysAlaLeuValHisGlnPheGlnProHisLysGly----- 112

Db 1636 GCAAAAGCCAGCCAAACAAATTAAGCAAGCAGAAAGACGCATAAAGGTGTAGTC 1695

QY 113 -----TyrGlyTyrPheAspTyrGlyLysAspIleAsnTrpGln 125

Db 1696 AAGGAGGCAAGCTTCAAGCTTATGGACATATTGAT----- 1731

QY 126 MetTrpProValLysAspAsnGluValArgTrpGlnLeuHisArgVal-----LysTrp 143

Db 1732 -----CAAGTTGATTTTGAGACAGGTGAAGTCTCTGGTAAATGG 1770

QY 144 -----TrpGlnAlaMetAlaLeuValTyrHisAlaThrGlyAspGluLysTyr 159

Db 1771 GATATTTTGTCTCGACTAGCTGTGTGTTAATAAGATTACCGGTGGAATTAACACG 1830

QY 160 AlaArgGluTrpValTyr-----GlnTyrSerAspTrpAlaArgLysAsnProLeuGly 177

Db 1831 GTTCTTGAATTTATGCATATCCCAACCATTCAGAGTGG-----AAGCCTAAAGGT 1881

QY 178 LeuSerGlnAspAsnAspLysPheValTrpArgProLeuGluValSerAspArgValGln 197

Db 1882 TACATGGCCGGTCTGAAAAAATGCAAAATGCGCCAGGAGCTGCATATCCGAAAGGTACA 1941

QY 198 SerLeuProProThrPheSerLeuPheValAsnSerPro-----AlaPheThr 213

Db 1942 GACTTTCCACCTGGAGGAAAGCGCTTGTGTGAAAAAGGATGGGAACCTGCCCATACA 2001

QY 214 ProAlaPheLeuMetGluPheLeuAsnSerTyrHisGlnGlnAlaAspTyrLeuSerThr 233

Db 2002 CCAGGCATCGGAATG----- 2016

QY 234 HisTyrAlaGluGlnGlyAsnHisArgLeuPheGluAlaGlnArgAsnLeuPheAlaGly 253

Db 2017 -----TATCTAGTCGT 2028

QY 254 ValSerPheProGluPheLysAspSerProArgTrpArgGlnThrGlyIleSerValLeu 273

Db 2029 ATGGCGCGCGCAAAATTTGGGACCTACCGGC-----GGCACATCTCTACTT 2076

QY 274 AsnThrGluIleLysGlnValTyrAlaAspGlyMetGlnPheGluLeuSerProIle 293

Db 2077 CCTCATGATCAGTCAAAAGAGCTTCAGCTTCGGGGCTT-----CTGGA 2121

QY 294 TyrHisValAlaAlaIleAspIlePheLeuLys---AlaTyrGlySerAlaLysArgVal 312

Db 2122 TATGCAGCGGGGTGCGGACTTTTTTAAAAAGCGCTGAAGGCTCTTAAGAAAAATGGTC 2181

QY 313 AsnLeuGluLysGluPheProGlnSerTyrValGlnThrValGluAsn----- 328

Db 2182 AATGGAGCTATTCTTTCGGAAGAGGTGTGTGGACAAAGTCTGGGGATGTTAGTTCAAGC 2241

QY 329 -----MetIleMetAlaLeuIleSerIleSer 337

Db 2242 GCTATGGACTGATATCAACGGCCCGCAAAATTAATTAATAAACTGTTAGCGGCTG 2301

QY 338 LeuProAspTyrAsnThr-----ProMetPheGlyAspSerTrpIleThrAsp 353

Db 2302 ATCCCG---TATAATCCGGAAGAGGTATGATTGTTGGAAACCGGCATCTCAAGACT 2358

QY 354 LysAsnPheArgMetAlaGlnPheAlaSerTrpAlaArgValPheProAla----- 370

Db 2359 TTAATAAATGTGTCAGCTCAATTTTTTGAAGAGAGTCATGCCAGAGCCGCTCCACATTTAA 2418

QY 371 -----AsnGlnAlaIleLysTyrPheAlaThrAsp-----GlyLysGlnGly 384

Db 2419 GGGACCGGAGAACGAAAGCGGTAAACCAATGGGTACAGAGGCTGTTGGTATACAGCGG 2478

QY 385 Lys-----AlaProAsnPheLeuSerLysAlaLeuSerAsnAlaGlyPheTyr 400

Db 2479 GTGCCACTTTCATGATTCAGGGCTGTGTGACCATCGGCATGAAGAAAGGGC----- 2532

QY 401 ThrPheArgSerGly-----TrpAspLysAsnAlaThrValMet 413

Db 2533 -----GGAAATCCAAATGCTATCAACTATATGGGATTCACATCGC----- 2571

QY 414 ValLeuLysAlaSerProProGlyGluPheHisAlaGlnProAspAsnGlyThrPheGlu 433
DB 2572 -----AAAGCAGGCCATCTTCACAGGGTTGATGCACAGCATCCGCGACACTTTT--- 2622
QY 434 LeuPheIleLysGlyArgAsnPheThrProAspAlaGlyValPheValTyrSerGlyAsp 453
DB 2622 ----- 2622
QY 454 GluAlaIleMetIysLeuArgAsnTrpTyrArgGlnThrArgIleHisSerThrLeuThr 473
DB 2622 ----- 2622
QY 474 LeuAspAsnGlnAsnMetValIleThrLysAlaArgGlnAsnLysTrpGluThrGlyAsn 493
DB 2623 -----AACGCCAATAAGTTCTCTGACATAAT 2649
QY 494 AsnLeuAspValLeuThrTyrThrAsnProSerTyrProAsnLeuAspHisGlnArgSer 513
DB 2650 AACATT-----TTGAATCG-----ATAGATAATAGTTGGCT 2682
QY 514 ValLeu-----PheIleAsnLysTyrPheLeuValIleAspArgAla----- 528
DB 2683 GCAATTAACTATATCAAGCGCATATGAGACATTAGCATCCAGGCTTAAAAATCA 2742
QY 529 -----IleGlyGluAlaThrGlyAsnLeu----- 536
DB 2743 ATGCCCGAGGTGGCGCTATGTGGTTATGCAAAAGGTGGTATCTCACCCTGACGGGGT 2802
QY 537 GlyValHisTrpGlnLeuLysGluAspSerAsnProValPheAspLysThrLysAsnArg 556
DB 2803 GGCTCAAAATGGCCATCTTAATGACGAGG-----TATGATGAACAACATCAGCAG 2856
QY 557 ValTyrThrTyrArgAspGlyAsnAsnLeuMetIleGlnSerLeuAsnAlaAspArg 576
DB 2857 GAAGATCCCTCAACAGGAGCGTAATATTGGGTATGGCTCGGATGGTGTAGCAAACTT 2916
QY 577 ThrSerLeuAsnGlu---GluGluGlyLysValSerTyrValTyrAsnLysGluLeuLys 595
DB 2917 GCGTACTCTCCGATTACAGGATGGGATGATTTCA-----AAAGCACTACTC 2964
QY 596 ArgProAlaPheValPheGluLysProLysLysAsnAlaGlyThrGlnAsnPheValSer 615
DB 2965 CTGCTCAAAAGTTTCGAAAGCCCAACAGCAGCTACCACTAATGATTTTCTGTA 3024
QY 616 IleValTyrPro-----TyrAspGlyGlnLysAla 625
DB 3025 GACATCAG-CCGTGTTCTGAAGAATCAAGAAAAACAATAACAGTATGATGGC----- 3074
QY 626 ProGluIleSerIleArgGluAsnLysGlyAsnAspPheGluLysGlyLys---LeuAsn 644
DB 3075 -----AAAGCAATTTGATTTTCTAAATAAAATGTCAGCT 3110
QY 645 LeuThrLeuThrIleAsnGlyLysGlnGln 654
DB 3111 CTTGCACGACTTTTACTAAAGACACAA 3140

RESULT 7

US-10-329-960-1/c
; Sequence 1, Application US/10329960
; Publication No. US20030099277A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Frag
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: P8186P1
; CURRENT APPLICATION NUMBER: US/10/329,960
; CURRENT FILING DATE: 2003-01-02
; PRIOR APPLICATION NUMBER: US 09/643,990
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 08/487,429
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787

; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (3921)..(3921)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (10150)..(10150)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (29298)..(29298)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36543)..(36543)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36636)..(36636)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (40808)..(40810)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44416)..(44416)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44905)..(44905)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44975)..(44975)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (45593)..(45593)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (45732)..(45732)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (47036)..(47036)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51334)..(51334)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51602)..(51602)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature

```
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (100091)..(100091)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (131360)..(131360)
```

```
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (152500)..(152500)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (152530)..(152530)
```

```
Alignment Scores:
Pred. No.: 590 Length: 1830121
Score: 115.00 Matches: 139
Percent Similarity: 32.80% Conservative: 85
Best Local Similarity: 20.35% Mismatches: 203
Query Match: 3.29% Indels: 256
DB: 15 Gaps: 41
```

US-09-802-285A-2 (1-659) x US-10-329-960-1 (1-1830121)

```
QY 34 PheAspHisIleAsnLeuGluTyrSerGlyLeuGluLysValAsnLysAlaValAlaLa 53
Db 676209 TGGGATCATATTAAACTAAGTTATTCTTCACAAAAAATCACCATAAAGCA----- 676159
QY 54 GlyAsnTyrAspAlaAlaLysAlaLeuLeuAlaTyrTyrArgGluLysSerLysAla 73
Db 676158 -----CGTTCTCATGAATATTGT 676141
QY 74 ArgGluProAspPheSerAsnAlaGluLysProAlaAspIleArgGlnProIleAspLys 93
Db 676140 CATCAATCCACTTGTAATGGTGAAGTAACCCCAAGGATTGCATTAGTAGAAGAAAA 676081
QY 94 ValThrArgGluMetAlaAspLys-----AlaLeu 103
Db 676080 GGGGTTTATAAAATTAAGGATAAATATGAGGTGAACITGAAAGCAAAAGAAATTGGCTGG 676021
QY 104 ValHisGlnPheGlnProHisGlyTyrGlyTyrPheAspTyrGlyLysAspIleAsn 123
Db 676020 TCTCATGAATTTAAGATAGTAGGGTGAA-----GATGCTGATAAGGATATTTC 675970
QY 124 TrpGlnMet-----TrpProValLysAspAsnGluValArg 135
:::
```



```
NAME/KEY: misc_feature
LOCATION: (9921)..(9921)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (10150)..(10150)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (29298)..(29298)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36543)..(36543)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36551)..(36551)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36636)..(36636)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (40808)..(40810)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44416)..(44416)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44905)..(44905)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44975)..(44975)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45593)..(45593)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45732)..(45732)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
```

```
LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (100091)..(100091)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (142750)..(142750)
```

```

/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (145058)..(145058)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (145171)..(145171)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (145942)..(145942)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (147197)..(147197)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (150841)..(150841)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (152500)..(152500)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (152530)..(152530)

```


Db 3025 GTCAACAAAGCAGAGTGGCTTCAACCGAAGAGTGTGATTCCGCAATAACAATGTTTATT 3084
QY 475 ---AspAsnGlnAsnMetValIleThrLysAlaArgGlnAsnLysTrpGluThrGly--- 492
Db 3085 GCAGAGAATTCATCACTACTAGT-ACT-----GATACATGGGGCTACGGGAAA 3131
QY 493 -----AsnAsnLeuAspValLeuThrTyrThrAsnPro 503
Db 3132 TTACCAACCCCATTTTTCAGCGAATAAAGAAATTCCTGTCGGTTTTTTTGGCGAATGC 3191
QY 504 SerTyrProAsnLeuAspHisGlnArgSerValLeuPheIleAsnLysLysTyrPhe--- 522
Db 3192 GGCTGGATAGATAAAGATGATCAGAACTTAATGATTTT-----AAGCGTAGTTTTATC 3245
QY 523 -----LeuValIleAspArgAlaIleGlyGluAlaThrGly 534
Db 3246 TACTTCAAGCCCATTAATTCAGTAATTAATTTATGATAGTTTATAGTCAAAAAGAGACT 3305
QY 535 AsnLeuGlyValHisTrpGlnLeuLysGluAspSerAsnProValPheAspLysThrLys 554
Db 3306 GAAATTACGAGTACTTATAATTTAGCACCTACGCTAAAT-----TGTCAAAAAAGCAGA 3359
QY 555 AsnArgValTyr-ThrThrTyrArgAspGlyAsnAsnLeuMetIleGlnSerLeuAsnAl 574
Db 3360 TCAATTTCAATTTAACTACAAATTCACAAAATATAAACTCCTTATAGTATCAGGTCAAAAC 3419
QY 574 aAspArgThrSerLeuAsnGluGluGluGlyLysValSerTyrValTyrAsnLysGluLe 594
Db 3420 GGATCAAACTACTGTGAAAGATCAGAG-----ATTATAATCAGCTGAT 3464
QY 594 uLysArgProAlaPheValPheGluLys-----ProLysLysAsnAl 608
Db 3465 CGAA-----CTATATACGCTTGACCAATAAATTTTCATTATAAAACGGGGAAGAGTTC 3518
QY 608 aGlyThrGlnAsnPheValSerIleValTyrPro---TyrAspGlyGlnLysAlaPro 626
Db 3519 AGGATACCGTGATTTCCCTTTGGAGAGATATCCAGATTATGCCAGTAAAGTGAACCA 3576

RESULT 11

US-09-993-292A-19
; Sequence 19, Application US/09993292A
; Patent No. US20020146430A1
; GENERAL INFORMATION:
; APPLICANT: James E. Galen
; TITLE OF INVENTION: USE OF CLY A HEMOLYSIN FOR EXCRETION OF
; TITLE OF INVENTION: PROTEINS
; FILE REFERENCE: UOFMD.007A
; CURRENT APPLICATION NUMBER: US/09/993,292A
; CURRENT FILING DATE: 2001-11-23
; PRIOR APPLICATION NUMBER: 60/252,516
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 2253
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ClyA::SacB fusion gene
; NAME/KEY: CDS
; LOCATION: (0)...(2253)
US-09-993-292A-19

Alignment Scores:
Pred. No.: 0.0501 Length: 2253
Score: 111.00 Matches: 124
Percent Similarity: 32.18% Conservative: 80
Best Local Similarity: 19.56% Mismatches: 196
Query Match: 3.18% Indels: 234
DB: 11 Gaps: 34

US-09-802-285A-2 (1-659) x US-09-993-292A-19 (1-2253)

QY 55 AsnTyrAspAsp-----AlaAlaLysAlaLeuLeuAlaTyr 66
Db 793 GATTATGATGATTAAATGCTTTCTTTTATAAAGAGAGCTGCAAGAAAATGATTAAACACC 852
QY 67 TyrArgGlu-----LysSerLysAlaArgGluProAspPheSer 79
Db 853 TGTATATGATACCAACAACGCTCATGGTAAGAGACGCTTTTCGAGGTTCTCGACGTCGCT 912
QY 80 AsnAlaGlu-----LysProAlaAspIleArgGlnProIleAspLysValThrArg 96
Db 913 AGTAAAGAAACCAACCAAAAGCCATATAGGAAACATACGCGCATTTCCCATATTATACGCG 972
QY 97 GluMetAlaAspLysAlaLeuValHisGlnPheGlnProHisLysGlyTyrGlyTyrPhe 116
Db 973 CAT-----GATATGCTGCAAAATCCTGAAACAGCAAAAATAATCAAAATATCAAGTTCCT 1026
QY 117 AspTyr-----GlyLysAspIleAsnTrp 124
Db 1027 GAATTCGATTGCTCCACAATTAATAATATCTCTTCTGCAAAAGGCGCTGGACGTT--TGG 1083
QY 125 GlnMetTrpProValLysAspAsnGluValArgTrpGlnLeuHisArgValLysTrpTrp 144
Db 1084 GACAGCTGGCCATTACAAAACGCTGAC-----GGCACTGTCGCAAACTAT 1128
QY 145 GlnAlaMetAlaLeuValTyrHisAlaThrGlyAspGluLysTyrAlaArgGluTrp--- 163
Db 1129 CAGCGCTACCATCGTCTTTGTCATTAGCCGAGATCCTTAAAAATGCGGATGACACATCG 1188
QY 164 ValTyrGlnTyr-----SerAspTrpAlaArgLysAsn 174
Db 1189 ATTTACATGTTCTATCAAAAGTCGCGCAAACTTCTATTGACAGCTGG-----AAAAAC 1242
QY 175 ProLeuGlyLeuSerGlnAspAsnAspLysPheValTrpArgProLeuGluValSerAsp 194
Db 1243 GCTGCGCGCGTCTTTAAAGACAGCGCAAAATTCGATGCAAAATGATTCTATCTCTAAAGAC 1302
QY 195 ArgValGlnSerLeuProThrPheSerLeuPheValAsnSerProAlaPheThrPro 214
Db 1303 CAACACACAGAA-----TGTTCAGGTTTCAGCCACATTTACATCT 1341
QY 215 AlaPheLeuMetGluPheLeuAsnSerTyrHisGlnGlnAlaAspTyrLeuSerThrHis 234
Db 1342 GACGGAAAAATCCGTTTATTC-----TACACTGATTCTCCGCTAAACAT 1386
QY 235 TyrAlaGluGlnGlyAsnHisArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyVal 254
Db 1387 TAC-----GGCAACAAACACTGCACTGCAAAAGTTAACGTTATCAGCATCA--- 1434
QY 255 SerPheProGluPheLysAspSerProArgTrpArgGlnThrGlyIleSerValLeuAsn 274
Db 1435 -----GACAGCTCTTGAACATCAACGCT 1458
QY 275 ThrGluIleLysGlnValTyr---AlaAspGlyMetGlnPheGluLeuSerProIle 293
Db 1459 GTAGAGGATTATAAATCAATCTTTCAGCGTGACGGA----- 1494
QY 294 TyrHisValAlaAlaIleAspIlePheLeuLysAlaTyrGlySerAlaLysArgValAsn 313
Db 1495 -----AAAAAGTATCAAAATGTACAGCAGTTTCATC 1524
QY 314 LeuGluLysGluPheProGlnSerTyrValGlnThrValGluAsnMetIleMetAlaLeu 333
Db 1525 GATGAAGGCAACTACAGCTCAGCGCACACCATACGCTGAGAGAT----- 1569
QY 334 IleSerIleSerLeuProAspTyrAsnThrProMetPheGlyAspSerTrpIleThrAsp 353
Db 1570 -----CCTCAGCTAC-----GTAGAAGAT 1587
QY 354 LysAsnPheArgMetAlaGlnPheAlaSerTrpAlaArgValPheProAlaAsnGlnAla 373
Db 1588 AAAGGCCACAAATACTTA-----GTATTTGAAGCAACACTGGA 1626


```
QY 294 TyrHisValAlaAlaIleAspIlePheLeuLysAlaTyrGlySerAlaLysArgValAsn 313
Db 2010 -----AAACGTAATCAAAATGACAGAGTTTCATC 2039
QY 314 LeuGluLysGluPheProGlnSerTyrValGlnThrValGluAsnMetIleMetAlaLeu 333
Db 2040 GATGAAGGCAACTACAGCTCAGCGCACCACTACGCTGAGAGAT-----2084
QY 334 IleSerIleSerLeuProAspTyrAsnThrProMetPheGlyAspSerTrpIleThrAsp 353
Db 2085 -----CCTCACTAC-----GTAGAAGAT 2102
QY 354 LysAsnPheArgMetAlaGlnPheAlaSerTrpAlaArgValPheProAlaAsnGlnAla 373
Db 2103 AAAGGCCACAATACTTA-----GTATTTCAAGCAAACTGGA 2141
QY 374 IleLysTyrPheAlaThrAspGlyLysGlnGlyLysAlaProAsnPheLeuSerLysAla 393
Db 2142 ACTGAA-----GATGGCTACCAAGCGAA-----GAATCT 2171
QY 394 LeuSerAsnAlaGlyPheTyrThrPheArgSerGlyTrpAspLysAsnAlaThrValMet 413
Db 2172 TTATTTACAAAGCATACTATGCGAAAGCACATCATCTCCGTCAGAAAGTCAAAA 2231
QY 414 ValLeuLysAlaSerProGlyGluPheHisAlaGlnProAspAsnGlyThrPheGlu 433
Db 2232 CTTCTGCAAAAGCGATAAA-----AAACGACGGCTGAGTAGCAACGGCGCTCTCGT 2285
QY 434 LeuPheIleLysGlyArgAsnPheThr-----442
Db 2286 ATGATGTAGCTAAACGATGATTACACATCGATGAAAAAGTGATGAAACCGCTGATTCATCT 2345
QY 443 -----ProAspAlaGlyValPheValTyrSerGlyAspGluAla 455
Db 2346 AACACAGTAACAGATGAATTCACCGCGCAAGCTTTAAATGAACGGC-----2396
QY 456 IleMetLysLeuArgAsnTrpTyrArgGlnThrArgIleHis---SerThrLeuThrLeu 474
Db 2397 -----AAATGGTACCTGTTCTAATCTTAACTGGCCCATACAGCG-----2513
QY 475 AspAsnGlnAsnMetValIleThrLysAlaArgGlnAsnLysTrpGluThrGlyAsnAsn 494
Db 2442 GACGGC-----ATTACGCTCAACGAT 2462
QY 495 LeuAspValLeuThrTyr-----ThrAsnProSerTyrProAsnLeuAsp 509
Db 2463 ATTTACATGCTTGGTTATGTTCTAATCTTAACTGGCCCATACAGCG-----2513
QY 510 HisGlnArgSerValLeuPheIleAsnLysLysTyrPheLeuValIleAspArgAlaIle 529
Db 2514 -----CTGAACAA-----2522
QY 530 GlyGluAlaThrGlyAsnLeuGlyValHisTrpGlnLeuLysGluAspSerAsnProVal 549
Db 2523 -----ACTGGCCTTGTG-----TTAAAAATGGATCTTGATCCT---2555
QY 550 PheAspLysThrLysAsnArgValTyrThrTyr-----Arg 562
Db 2556 -----AACGATGTAACCTTACTTACTACACTTCGCTGTACCTCAAGCG 2600
QY 563 AspGlyAsnAsnLeuMetIleGlnSer-----LeuAsnAlaAspArg 576
Db 2601 AAAGGAAACAATGTCGTGATTACAGCTATATGACAAACAGAGGATTTCTACGCACACAA 2660
QY 577 ThrSer-----LeuAsnGluGluGluGlyLysValSerTyrVal 589
Db 2661 CAATCAACGTTTGGCCCAAGCTTCTGCTGAAACATCAAGGCAAGAAACATCTCTTGTCT 2720
QY 590 TyrAsnLysGluLeuLysArgProAlaPheValPheGluLys 603
Db 2721 AAAGACAGCATCTTGAACAAAGGACAAATTAACAGTAACAAA 2762
```

RESULT 13

```
US-09-815-242-4610
; Sequence 4610, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4610
; LENGTH: 1689
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-4610

Alignment Scores:
Pred. No.: 0.0365 Length: 1689
Score: 110.50 Matches: 105
Percent Similarity: 34.25% Conservative: 94
Best Local Similarity: 18.07% Mismatches: 211
Query Match: 3.16% Indels: 171
DB: 10 Gaps: 26

US-09-802-285A-2 (1-659) x US-09-815-242-4610 (1-1689)
QY 91 IleAspLysValThrArgGluMetAlaAspLysAlaLeuValHisGlnPheGlnProHis 110
Db 22 ATCGNACCCATTTTAATGAATATTTAGAGAAGAAATTTATATCAAGTACCCTCAA 81
QY 111 LysGlyTyrGlyTyrPheAspTyrGlyLysAspIleAsnTrpGlnMetTrpProValLys 130
Db 82 AAT-----CATCAATGGTCTATACAA 102
QY 131 AspAsnGluValArgTrpGlnLeuHisArgValLysTrpTrpGlnAlaMetAla---Leu 149
Db 103 TTATCTGAGCTCGAAACACGTTAACAGGTCATTCGCTTATTCGTCGATGGGCGATCAT 162
QY 150 ValTyrHisAlaThrGly-----AspGluLysTyrAlaArgGluTrpValTyrGln 166
Db 163 ATGTATCATTCAGAGTATGCTTATTCATGTTAAAGTAAGAAATTAACAACATATAA 222
QY 167 TyrSerAspTrpAlaArgLysAsnProLeuGlyLeuSerGlnAspAsnLysPheVal 186
Db 223 GAAGCCATTGGCGGTATTTTCGAACATATGCTCAAGTGCAGATAATCAACACGCGATG 282
QY 187 TrpArgProLeuGluValSerAspArgValGlnSerLeuProProThrPheSerLeuPhe 206
Db 283 CAACAACTATG-----294
```

```
QY 207 ValAsnSerProAlaPheThrProAlaPheLeuMetGluPheLeuAsnSerTyrHisGln 226
Db 295 -----GGCAGAAATATGCTGACATCGATATAGCATTCATCGC 333
QY 227 GlnAlaAspTyrLeuSerThrHisTyrAlaGluGlnGlyAsnHisArgLeuPheGluAla 246
Db 334 ACGCGCGGTATTTGCAAGATACACAATAGATACCGAGAGATCGCTATATCGTTTCA 333
QY 247 GlnArgAsnLeuPheAlaGlyValSerPhe---ProGluPheLysAspSerProArgTrp 265
Db 394 GAACAATCTTTATACCTAGGTATCCATTTTCATCCGACTCCTAAGAGTGCAGTGGGTT 453
QY 266 ArgGlnThrGlyIleSer----- 271
Db 454 TCGAAGCAGATTTAGAGAAATATGACCGCAATGTCATACATCATTCCAATTCGATTAT 513
QY 272 ---ValLeuAsnThrGluIleLysLysGlnValTyrAlaAspGlyMetGlnPheGluLeu 290
Db 514 TTAGCTGTGCATCAAGATGTTCTGCTCAGCGCTATGTAGAGGTAAAGAAGATCAGTT 573
QY 291 SerProIle---TyrHisValAlaAlaIleAspIle-----PheLeu 303
Db 574 GAGAAAGTGTGTATCAATTTAGCAGATATATAGCATATCAGAGATACCTAAAGATTTTAT 633
QY 304 LysAlaTyrGlySerAlaLysArgValAsnLeuGluLysGluPheProGlnSerTyrVal 323
Db 634 TTATTACCAATACATCCTTATCAAAATGATGTGTCACAGCATCCACAG---TATTG 690
QY 324 GlnThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuProAspTyrAsnThr 343
Db 691 CAATATAGTGAACAAGGTTTAAATAAAGACCTTGGCGTTCC----- 732
QY 344 ProMetPheGlyAspSer----- 349
Db 733 -----GGTGATTTCAGTTTACCGACGCTCTCGGTTAGAACTGTATTTTCAAAAGCA 783
QY 350 -----TrrPheThrAspLysAsnPhe---ArgMetAla 359
Db 784 TTAACATTTTATTAATAATACCGATACAGCTTAAATTAATAATTTATACGTACGAAT 843
QY 360 GlnPheAlaSerTrpAlaArgValPheProAlaAsnGlnAlaIle----- 374
Db 844 GATCTTGAACAGATTTGAACGCAATCGATGCTGCACAAGTTATCGCATCAGTCAAAAGAT 903
QY 375 -----LysTyrPheAlaThrAspGlyLysGlnGlyLysAlaPro 387
Db 904 GAGGTGAAACACCCCATTTTAAATGATGTTTGAAGAAGGATATCGTGCATTTGTACCG 963
QY 388 AsnPheLeuSerLysAla-----LeuSerAsnAlaGlyPheTyr 400
Db 964 AATCCATTAGGGCAACAGTTGACCTGAATGATTTATTAACAATAGTGCATGATT 1023
QY 401 ThrPheArgSerGly-----TrpAspLysAsnAlaThrValMetValLeu 415
Db 1024 GTT---CGTGAAGGATACCGAATTACCATGCTGATAAAGATATTCATGTATTGGCG--- 1077
QY 416 LysAlaSerProGlyGluPheHisAlaGlnProAspAsnGlyThrPheGluLeu--- 434
Db 1078 -----TCATTATTGAACCGATGCTGATTACCGACCTCTAAGATTATCA 1122
QY 435 ---PheIleLysGlyArgAsnPheThrProAspAlaGlyValPheValTyrSerGlyAsp 453
Db 1123 CAAGTATTGAGCAAACTGGTTAGCCCGCAGACATGGCTTGNATGTTATTG---GAT 1179
QY 454 GluAlaIleMetLysLeuArgAsnTrpTyrArgGlnThrArgIle----- 468
Db 1180 CGTACATTATTGCCGATATTAAAGCTGTATAGTAAACACAGCAATTAGTCTAGAAGCACAT 1239
QY 469 -----HisSerThrLeuThrLeu-----Asp 475
Db 1240 GTACAAAATTCATTAATTAATGAATTAATAAAGATGGCATACCCGACGATGCTTTGTGACAGAT 1299
QY 476 AsnGlnAsnMetValIleThrLysAlaArgGlnAsnLysTrpGluThrGlyAsnAsnLeu 495
```

```
Db 1300 CTTGAAGGCATTTGCTATCTAGAACGATTGCTACTGAAAAACAATCTGTGCCAAAT--- 1356
QY 496 AspValLeuThrTyrThrAsnProSerTyrProAsnLeuAspHisGlnArgSerValLeu 515
Db 1357 ---GTTGTGGCAGCATCTAGCCCT-----GTTGTATATGCACATGATGAAGCG 1401
QY 516 PheIleAsnLysLysTyrPheLeuValIleAspArg-----AlaIle 529
Db 1402 TGCCATCGCTTAAATATTACGTTGTAGTAAATCACTTAGGACATTTAGTATCACTATT 1461
QY 530 GlyGluAlaThrGlyValHisTyrGlnLeu----- 542
Db 1462 GGTAAAGCACAAGATGAAGTTGTGTATGGCACTTGTAGCGCATCGTCTTATGACT 1521
QY 543 -----LysGlu---AspSerAsnProValPheAspLysThrLysAsnArgValTyrThr 559
Db 1522 TGAATAAAGAAATACGCAATACGCGATATTGTTGACTGTGTAGAAAGATTTATATCAA 1581
QY 560 Thr 560
Db 1582 ACG 1584

RESULT 14
US-09-815-242-8108
; Sequence 8108 Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8108
; LENGTH: 1737
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1737)
US-09-815-242-8108

Alignment Scores:
Pred. No.: 0.0381 Length: 1737
Score: 110.50 Matches: 105
Percent Similarity: 34.25% Conservative: 94
Best Local Similarity: 18.07% Mismatches: 211
Query Match: 3.16% Indels: 171
DB: 10 Gaps: 26
```

US-09-802-285A-2 (1-659) x US-09-815-242-8108 (1-1737)

```
QY 91 IleAspLysValThrArgGluMetAlaAspLysAlaLeuValHisGlnPheGlnProHis 110
Db 43 ATCGAACGCAATTAATGAATATTTAGAGAGAAAATTTATATCAAGTACCACCTCAA 102
QY 111 LysGlyTyrGlyTyrPheAspTyrGlyLysAspIleAsnTrpGlnMetTrpProValLys 130
Db 103 AAT-----CATCAATGGTCTATACAA 123
QY 131 AspAsnGluValArgTrpGlnLeuHisArgValLysTrpGlnAlaMetAla---Leu 149
Db 124 TTATCTGAGCTCGAAACGTTAAACAGGTCAATCGCTTATTTGGTCTCGGATGGGCATCAT 183
QY 150 ValTyrHisAlaThrGly-----AspGluLysTyrAlaArgGluTrpValTyrGln 166
Db 184 ATGTATCATCAAGATATGGCTTATGATGATGTAAGAAAGTAAAGAAATTAACAACATATAAA 243
QY 167 TyrSerAspTrpAlaArgLysAsnProLeuGlyLeuSerGlnAspAsnAspLysPheVal 186
Db 244 GAAGCAATGGCGGTATTTTGCACATATGGCTCAAAGTGCAGATTAATCAACCGCAGTG 303
QY 187 TrpArgProLeuGluValSerAspArgValGlnSerLeuProProThrPheSerLeuPhe 206
Db 304 CAACAACATATG----- 315
QY 207 ValAsnSerProAlaPheThrProAlaPheLeuMetGluPheLeuAsnSerTyrHisGln 226
Db 316 -----CGCAAAATATGCTGACATCAATATAGCATTCATCGC 354
QY 227 GlnAlaAspTyrLeuSerThrHisTyrAlaGluGlnGlyAsnHisArgLeuPheGluAla 246
Db 355 ACGCGCGGTATTTGCAAGTAACACATAGACTACGCAGAGATCGCTATATCGTTTCA 414
QY 247 GlnArgAsnLeuPheAlaGlyValSerPhe---ProGluPheLysAspSerProArgTrp 265
Db 415 GAACAATCTTTATCTTAGTCTATCCATTTTCATCGACTCCTTAAGAGTGAAGTGGGTTT 474
QY 266 ArgGlnThrGlyLeuSer----- 271
Db 475 TCAGAAGCAGATTTAGAGAAATATGACCCGAAATGTACATCATCTCCAAATGGCATAT 534
QY 272 ---ValLeuAsnThrGluIleLysLysGlnValTyrAlaAspGlyMetGlnPheGluLeu 290
Db 535 TTAGCTGTGTCATCAAGATTTCTGCTCACGCGCTATGTAGAGGTAAAGAGATCAGGTT 594
QY 291 SerProIle---TyrHisValAlaAlaIleAspIle-----PheLeu 303
Db 595 GAGAAAGTGTGTATCAATATAGCATATAGACATATAGACATATCAGAGATACCTTAAGATTTTAT 654
QY 304 LysAlaTyrGlySerAlaLysArgValAsnLeuGluLysGluPheProGlnSerTyrVal 323
Db 655 TTATTACCAATACATCTTATCAATCAATGTTGTCACACATCCACAG---TATTTG 711
QY 324 GlnThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuProAspTyrAsnThr 343
Db 712 CAATATAGTGAACAAGGTTTATATAAAGACCTTGGCGTTTC----- 753
QY 344 ProMetPheGlyAspSer----- 349
Db 754 -----GGTGATTCAGTTTACCAGCTCTCCGTTAGAACTGTATTTTCAAAGCA 804
QY 350 -----TrpIleThrAspLysAsnPhe---ArgMetAla 359
Db 805 TTAAACATTTATTAATACCGATACAGTGTAAATTTACTAAATTTATACGTACGAT 864
QY 360 GlnPheAlaSerTrpAlaArgValPheProAlaAsnGlnAlaIle----- 374
Db 865 GATCTTGAACAGATTGAACGGCAATCGATGTCGACAAAGTTATCGCATCATCAAGAT 924
QY 375 -----LysTyrPheAlaThrAspGlyLysGlnGlyLysAlaPro 387
```

```
Db 925 GAGGTTGAAACACCCCATTTTAAATGTATGTTTGAAGAAGGATATCGTGCATTTGTACCG 984
QY 388 AsnPheLeuSerLysAla-----LeuSerAsnAlaGlyPheTyr 400
Db 985 AATCCATTAGGGCAACACAGCTGAACCTGAAATGGATTTATTATTAACAATAGTGCATGATT 1044
QY 401 ThrPheArgSerGly-----TrpAspLysAsnAlaThrValMetValLeu 415
Db 1045 GTT---CGTGAAGGATACCCGAATTAACCATGCTGATAAAGATATTCATGTATTGGCG--- 1098
QY 416 LysAlaSerProGlyGluPheHisAlaGlnProAspAsnGlyThrPheGluLeu--- 434
Db 1099 -----TCATTATTGAAACGATGCTGATTACCGACCTCTAAGTTATCA 1143
QY 435 ---PheIleLysGlyArgAsnPheThrProAspAlaGlyValPheValTyrSerGlyAsp 453
Db 1144 CAAGTGATTGACCAAGATGGTTTAGCGCCAGAGCATGGCTTCAATGTTATTG---GAT 1200
QY 454 GluAlaIleMetLysLeuArgAsnTrpTyrArgGlnThrArgIle----- 468
Db 1201 CGTACATATTGCCGATATAAAGCTGTATAGTAAACACAGGCATTAGCTCTAGAAGCACAT 1260
QY 469 ---HisSerThrLeuThrLeu-----Asp 475
Db 1261 GTACAAAATTCATTAAATGAAATTAAGATGGCATACCGACGTATGCTTTGTCAGAGAT 1320
QY 476 AsnGlnAsnMetValIleThrLysAlaArgGlnAsnLysTrpGluThrGlyAsnLeu 495
Db 1321 CTTGAAGGCATTTGTCTATCTAGAACGATGCTACTGAAAAACAACACTTGTGCCAAAT--- 1377
QY 496 AspValLeuThrTyrThrAsnProSerTyrProAsnLeuAspHisGlnArgSerValLeu 515
Db 1378 ---GTTGGCGCAGCATCTAGCCCT-----GTTGTATATGCACATGATGAAGCG 1422
QY 516 PheIleAsnLysLysTyrPheLeuValIleAspArg-----AlaIle 529
Db 1423 TGGCATCGCTCTAAATATATACGTTCTAGTAAATCACTTAGGACATTTAGTATCAACTATT 1482
QY 530 GlyGluAlaThrGlyAsnLeuGlyValHisTyrGlnLeu----- 542
Db 1483 GGTAAAGGCACAAAGATGAAGTTGTATTATGGCAACTGTAGCCGATCGTCTTATGACT 1542
QY 543 -----LysGlu---AspSerAsnProValPheAspLysThrLysAsnArgValTyrThr 559
Db 1543 TGGAAAAAAGATACACGCGAATACCCAGATTTTGTGACTGTGTAGACATTTATATCAAA 1602
QY 560 Thr 560
Db 1603 ACG 1605
```

RESULT 15

US-09-070-927A-91/C

; Sequence 91, Application US/09070927A

; Patent No. US20020120116A1

; GENERAL INFORMATION:

APPLICANT: Charles A. Kunsch

Steven Barash

Patrick J. Dillon

TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides

NUMBER OF SEQUENCES: 982

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII TEXT

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/070,927A
 FILING DATE: 04-May-2000
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 60/046,655
 FILING DATE: 1997-05-16
 APPLICATION NUMBER: 60/044,031
 FILING DATE: 1997-05-06
 APPLICATION NUMBER: 60/066,009
 FILING DATE: 1997-11-14
 ATTORNEY/AGENT INFORMATION:
 NAME: Kenley K. Hoover
 REGISTRATION NUMBER: 40,302
 REFERENCE/DOCKET NUMBER: PB369
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 309-8504
 TELEFAX: (301) 309-8512
 INFORMATION FOR SEQ ID NO: 91:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 10194 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 SEQUENCE DESCRIPTION: SEQ ID NO: 91:
 US-09-070-927A-91

Alignment Scores:

Pred. No.: 0.596 Length: 10194
 Score: 110.50 Matches: 143
 Percent Similarity: 36.54% Conservative: 115
 Best Local Similarity: 20.25% Mismatches: 274
 Query Match: 3.16% Indels: 178
 DB: 11 Gaps: 32

US-09-802-285A-2 (1-659) x US-09-070-927A-91 (1-10194)

```

15 IleAlaLeuSerSerGlyAsnIleLeuAlaGlnSerSerSer----- 28
9401 GTATCTCTTAAAGGAGGAGATATCTTCACATTCGAAACTGCCGACAAAGATTTTGTGTT 9342
29 -----IleThrArgLysAspPheAspHisIle-----AsnLeuGlu 40
9341 AGATTTTAAATGAAAAAGATTTTCAAGAATATATGAAAAATCCAGTATCMAGTACTGAA 9282
41 TyrSerGlyLeuGluLysValAlaLeuLysAlaValAlaAlaGlyAsnTyrAspAlaAla 60
9281 ACTAGCAACAAACAAACAGTAAACTCTGTATGTTTCTAAAGTAGTAGCCCAAGATAATAA 9222
61 LysAlaLeuLeuAlaTyrTyrArgGluLysSerLysAlaArgGluProAspPheSerAsn 80
9221 CAATCTGATGTAATCTGAAAAAAGAGAGTAAGTACTCAAGCGAAGTCTGATGTAGTACT 9162
81 AlaGluLysProLa-----AspIleArgGlnProIleAspLysVal 94
9161 AATACTTTTACCGAGCGAAGATAAAATACTAATGACATTACTAAGTAGTAGCAGATGAGCCA 9102
95 ThrArgGlu-----MetAlaAspLysAlaLeuValHisGlnPheGln----- 108
9101 ACCTTAGAACAAACACCGCTTAGATACCTTTAGTAGTACATCAATTTAATGATATGTAT 9042
109 ProHisLysGlyTyrGlyTyrPheAspTyrGly-LysAspIleAsnTrpGlnMetTrpPr 128
9041 CCTTATAAGGAACAAATG-----CATTCATATTCGGCGTCATCCACCATGGAC 8988
128 oValLysAspAsnGluValArgTrpGlnLeuHisArgValLysTrpGlnAlaMet-- 147
8987 GCAAAAAGATGTT-----AAATGGTATCAAAAAGTAG 8955
148 -AlaLeuValTyrHisAlaThrGlyAspGluLysTyrAlaArgGluTrpValTyr----- 165
8954 TGCTACTATAGTAATATGATATGATGAGCAAAAGAGAGAGCTAATG--TGGAAATACACATA 8897
166 -----GlnTyrSerAspTrp----- 170

```

```

8896 ACTCCACAAGTCAGATAGCGAGCTGTAGAAATTTATGATTTATTAATCTTCTTATT 8837
171 -----AlaArgLysAsnProLeuGlyLeuSerGlnAspAsnAspLysPheValTr 187
8836 ATTTTATTAAACAAAGAGAGAAACCTTCGGGCTTTTCTTTTATACAAAAACGAACATA 8777
187 pArgProLeuGluValSerAspArgValGln-SerLeuProProThrPheSerLeuPheV 207
8776 TATTCGTATTTCGTATTATGAT--ATACGATACTTGATCTTACTTATAACATATTC- 8721
207 alAsnSerProAlaPheThrProAlaPheLeuMetGluPheLeuAsnSerTyrHisGlnG 227
8720 --AATAAAAAATCGTCTTAAATGATTTAAATTA----- 8688
227 lnAlaAspTyrLeuSerThrHisTyrAlaGlnGlnGlyAsnHisArgLeuPheGluAlaG 247
8687 -----AAAAATCCGAGAGGTTGGTATTATTCGCAG 8657
247 lnArgAsnLeuPheAlaGlyValSerPhePro-----GluPheLysAspSer-- 262
8656 AACCAAGAAATTAGCTAATGGAATGCGAGTTAAGATTAAATATAAAGACTCTATTA 8597
263 --ProArgTrpArgGlnThrGlyLysSer-----ValLeuA 274
8596 CAATCAATGGAATAAATAAATCAATCTACTAGATCTACTAAGAGGCTGTAGAGATGCTG 8537
274 snThrGluLeuLysLysGlnValTyr-----AlaAspGlyMetGlnPheGluLeuSerP 292
8536 AAACAGAGTTTAAATCAAAATATGCGTGGTGAACAACTGAAGCAATTTAAACTTTTAG 8477
292 roIleTyrHisValAlaAlaIleAspIlePheLeuLysAlaTyrGlySerAlaLysArgV 312
8476 ATTTTATGATATT--TGGTTGATACATTTAAAGAATAAAGTGTCTGCTGGCGGT- 8421
312 alAsnLeuGluLysGluPheProGlnSerTyrValGlnThrValGluAsnMetIleMetA 332
8420 -----ATGCAAAA-----ATAG 8408
332 laLeuIleSerIleSerLeuProAspTyrAsnThrProMetPheGlyAspSerTrpIleT 352
8407 CTTTAAACAAGAAAGATTTTACATGATTC-----TTTGGGACAAACAATTCG 8360
352 hrAspLysAsnPheArgMetAlaGlnPheAlaSerTrpAlaArgValPheProAlaAsnG 372
8359 TA---AAAGCGGTCAACAAAGTCAAAATATCAGCAATG----- 8325
372 lnAlaIleLysTyrPheAlaThrAspGly-----LysGlnGlyLysAlaProAsnP 389
8324 -----ATTAACTGGCTTGCAAAACCTGGCAATATCAATGAAAAGAGCTAGCAGTTGAG 8270
389 heLeuSerLysAlaLeuSerAsnAlaGlyPheTyrThrPheArgSerGlyTrpAspLysA 409
8269 CGGTGAGCAACAGACATAA-CATTGTAAATCTATCTTTTAAAGCGT-TGCACATGCA 8212
409 snAlaThrValMetValLeuLysAla-----SerProGlyGluPheHisAlaGlnP 427
8211 ATATATTCTTCAATCCAAACCAAGAAATATAAGTTA-ACTGGCCCAAGTACCAAAAAGA 8153
427 roAspAsnGlyThrPhe-----GluLeuPheIleLysG 438
8152 AATCAATATAAACCATATCTATTGATGATTTAACTAAATTTAAGAAAGCTTTATTATCTC 8093
438 lyArgAsnPheThrProAspAlaGlyValPheVal-----TyrSerGlyAspGluAla 456
8092 GTGAAAACACCAACAGTAAATATCTCTTTTAGTACAACTATATACTGAGCTCGT---- 8037
456 leMetLysLeuArgAsnTrpTyrArgGlnThrArgIleHisSerThrLeuThrLeuAspA 476
8036 -----TATCAAGAGGTTCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7997
476 snGlnAsnMetValIleThrLysAlaArgGlnAlaLysTrpGluThrGlyAsnAsnLeuA 496

```

```
Db 7996 AAAGTAACGAGCTTATTAAAGATAAAAAATGCCTATAAGTATGATGCTGCCAGTATAGAA 7937
QY 496 spValLeuThrTyrThrAsnProSerTyrProAsnLeuAspHisGlnArgSerValLeuP 516
Db 7936 TTGGCCCTACAAAATCAGAACGAGGTTCGAAATATTGATGTCCTTCTCTTTCT 7877
QY 516 heileAsnLysLysTyrPheLeuValIleAspArgAlaIle-----GlyGluAlaThrG 534
Db 7876 ATTATTAAAAAAATGAAATGGACAAAAGAAAAAATATTGACAGAGATTTACGTA 7817
QY 534 lyAsnLeuGlyValHisTrpGlnLeuLysGluAspSerAsnProValPheAspLysThrL 554
Db 7816 ATCCCCAAAATTTTGCTTTTGTAGTAAAAAAATTCATGGCCCATCTCAAATTCATCTG 7757
QY 554 ysAsnArgValTyrThr-----T 560
Db 7756 TGAATAAATATATATAAAGAAACATGTGAACCTAGCAAAATATTGAAAGAAATTCATCTCATA 7697
QY 560 hrTyrArgAspGlyAsnAsnLeuMetIleGlnSerLeuAsnAlaAsp-----A 576
Db 7696 GTTTTAGACATGCCCGTTCTGACTTATTATTCCTAGCTGAAGCTGATCCAATATATATCA 7637
QY 576 rgThrSerLeuAsnGluGluGlyLysValSerTyrValTyr-AsnLysGluLeuLys 595
Db 7636 AAACACAGCTAGCACATAAAGAGATAATACAGCTCATATGATACCGCATCAGCTACTGAAG 7577
QY 596 ArgProAlaPheValPheGluLysProLysLysAsnAlaGlyThrGlnAsnPheValSer 615
Db 7576 CAA-----ATAGAAAAAGACAAAGAAAAATATGAAAAACAAATACAAAGATATTT 7526
QY 616 ileValTyrPro 619
Db 7525 TGTAATAATACCCA 7514
```

Search completed: August 4, 2003, 13:41:05
Job time : 4172 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 4, 2003, 10:13:19 ; Search time 85 Seconds
(without alignments)

2377.647 Million cell updates/sec

Title: US-09-802-285a-2

Perfect score: 3494

Sequence: 1 MTKIFKRIIVFAVIALSSG.....KGLNLTLTKNGQQLVLPV 659

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame.p2n.model -DRV=xlh
-Q=/cgn2_1/USPTO.spool/US0802285/runat_28072003_190513_9740/app_query.fasta_1.839
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptlo -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0802285.@cgn2_1_40 -runat_28072003_190513_9740 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3494	100.0	1980	1	US-08-258-639A-3
2	3494	100.0	1980	2	US-08-900-951-3
3	3494	100.0	1980	5	PCT-US95-07391A-3
C	1276	5.0	3763	4	US-08-961-527-186
5	132.5	3.8	2097	4	US-09-134-001C-795
6	123.5	3.5	3414	1	US-07-973-320-1
7	123	3.5	1542	1	US-08-328-962-1
C	120.5	3.4	4403765	4	US-09-103-840A-2
C	120.5	3.4	4411529	4	US-09-103-840A-1
10	115.5	3.3	3414	1	US-07-973-320-3
11	115.5	3.3	3504	1	US-08-620-717A-8
12	115	3.3	2859	5	PCT-US96-05320A-637

13	115	3.3	16535	4	US-08-961-527-74	Sequence 74, Appl
14	114.5	3.3	3504	1	US-08-485-568A-5	Sequence 5, Appl
15	114.5	3.3	3504	2	US-08-590-554A-5	Sequence 5, Appl
16	114.5	3.3	3504	2	US-09-184-223-5	Sequence 5, Appl
17	113.5	3.2	4765	1	US-08-750-532-8	Sequence 8, Appl
18	113.5	3.2	4765	4	US-08-894-818B-7	Sequence 7, Appl
19	113.5	3.2	4765	4	US-09-445-472-5	Sequence 5, Appl
20	112.5	3.2	2339	1	US-08-258-639A-1	Sequence 1, Appl
21	112.5	3.2	2339	1	US-08-900-951-1	Sequence 1, Appl
22	112.5	3.2	2339	5	PCT-US95-07391A-1	Sequence 1, Appl
23	111	3.2	2784	4	US-09-134-001C-1994	Sequence 1994, Ap
24	109	3.1	2858	3	US-08-816-346-3	Sequence 3, Appl
25	109	3.1	2858	3	US-08-816-346-57	Sequence 57, Appl
26	109	3.1	2858	3	US-09-335-411-3	Sequence 3, Appl
27	109	3.1	2858	3	US-09-335-411-57	Sequence 57, Appl
28	109	3.1	2859	2	US-08-788-674-7	Sequence 7, Appl
29	109	3.1	34303	2	US-08-735-609-4	Sequence 4, Appl
30	109	3.1	34303	2	US-08-735-609-4	Sequence 4, Appl
31	109	3.1	34303	3	US-09-315-372-4	Sequence 4, Appl
32	109	3.1	34303	3	US-09-244-752-4	Sequence 4, Appl
33	109	3.1	34303	3	US-09-245-497-4	Sequence 4, Appl
34	109	3.1	34303	4	US-09-562-919-4	Sequence 4, Appl
35	109	3.1	34382	2	US-08-374-483-6	Sequence 6, Appl
36	109	3.1	35408	4	US-08-973-334-3	Sequence 3, Appl
37	109	3.1	35408	4	US-09-563-869A-3	Sequence 3, Appl
38	109	3.1	35408	4	US-08-549-489-3	Sequence 3, Appl
39	109	3.1	35935	2	US-08-735-609-1	Sequence 1, Appl
40	109	3.1	35935	2	US-08-735-609-1	Sequence 1, Appl
41	109	3.1	35935	3	US-08-379-452-43	Sequence 43, Appl
42	109	3.1	35935	3	US-09-315-372-1	Sequence 1, Appl
43	109	3.1	35935	3	US-09-244-752-1	Sequence 1, Appl
44	109	3.1	35935	3	US-09-245-497-1	Sequence 1, Appl
45	109	3.1	35935	3	US-09-409-670-43	Sequence 43, Appl

ALIGNMENTS

RESULT 1

US-08-258-639A-3
; Sequence 3, Application US/08258639A

; Patent No. 5681733

; GENERAL INFORMATION:

; APPLICANT: Su, Hongsheng

; APPLICANT: Blain, Francoise

; APPLICANT: Bennett, Clark

; APPLICANT: Gu, Kangfu

; APPLICANT: Zimmermann, Joseph

; APPLICANT: Musil, Roy

; TITLE OF INVENTION: Nucleic Acid Sequences And Expression

; TITLE OF INVENTION: Systems For Heparinase II And Heparinase III Derived From

; TITLE OF INVENTION: Flavobacterium heparinum

; NUMBER OF SEQUENCES: 26

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Hale and Dorr

; STREET: 1455 Pennsylvania Avenue, N.W.

; CITY: Washington, D.C.

; COUNTRY: U.S.A.

; ZIP: 20004

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/258,639A

; FILING DATE: 10 JUNE 1994

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Healey, William J.

; REGISTRATION NUMBER: 36,160

; REFERENCE/DOCKET NUMBER: 104385.116

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202)942-8400

TELEFAX: (202) 942-8484
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1980 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-258-639A-3

Alignment Scores:

Pred. No.: 0 Length: 1980
Score: 3494.00 Matches: 659
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-09-802-285A-2 (1-659) x US-08-258-639A-3 (1-1980)

QY 1 MetThrThrLysLysPheLysArgLleLleValPheAlaValIleAlaLeuSerSerGly 20
DB 1 ATGACTACGAAATTTTAAAGGATCATGTGATTTGCTGTAAATGCCCTATCGTCGGGA 60
QY 21 AsnIleLeuAlaGlnSerSerLleThrArgLysPheAspHisIleAsnLeuGlu 40
DB 61 AATATACTTGGCAAAAGCTCTCCATTACCAGGAAAGATTTTGACCAATCAACCTTGAG 120
QY 41 TyrSerGlyLeuGluLysValAsnLysAlaValAlaAlaGlyAsnTyrAspAlaAla 60
DB 121 TATTCGGACTGGAAAGGTTAATAAAGCAGTTGCTCGCGCAACTATGACGATGCGGCC 180
QY 61 LysAlaLeuLeuAlaTyrTyrArgGluLysSerLysAlaArgGluProAspPheSerAsn 80
DB 181 AAAGCACTACTGGCATACTACAGGGAAGAAAGTAAGCCAGGAACTGATTCAGTAAT 240
QY 81 AlaGluLysProAlaAspIleArgGlnProIleAspLysValThrArgGluMetAlaAsp 100
DB 241 GCAGAAAGCCCTGCCGATATAGCCAGCCATAGATAAGGTTACGGTGAATGGCCGAC 300
QY 101 LysAlaLeuValHisGlnPheGlnProHisLysGlyTyrGlyTyrPheAspTyrGlyLys 120
DB 301 AAGGCTTTGGTCCACCAGTTTCAACCGCAACAAGGCTACGGCTATTGTTGATTATGGTAAA 360
QY 121 AspIleAsnTrpGlnMetTrpProValLysAspAsnGluValArgTrpGlnLeuHisArg 140
DB 361 GACATCAACTGGCAGATGTGGCGGTAAAGACATGAAGTACGCTGGCAGTTGCACCT 420
QY 141 ValLysTrpTrpGlnAlaMetAlaLeuValTyrHisAlaThrGlyAspGluLysTyrAla 160
DB 421 GTAAAAATGGTGGCAGGCTATGGCCCTGGTTTATCACGCTACGGGCGATGAAAAATATGCA 480
QY 161 ArgGluTrpValTyrGlnTyrSerAspTrpAlaArgLysAsnProLeuGlyLeuSerGln 180
DB 481 AGAGAAATGGGTATATCAGTACAGCGAATGGGCCAGAAAAAACCATTGGCGCTGTCCGAC 540
QY 181 AspAsnAspLysPheValTrpArgProLeuGluValSerAspArgValGlnSerLeuPro 200
DB 541 GATAATGATAAATTTGTGTGGCGGCCCTTGAGTGTGCGACAGGTTACAAAGTCTTCCTCC 600
QY 201 ProThrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGluPhe 220
DB 601 CCAACCTTCAGCTTATTTGTAACTCGCCAGGCTTTTACCCAGCCCTTTTAAATGGAATTT 660
QY 221 LeuAsnSerTyrHisGlnGlnAlaAspTyrLeuSerThrHisTyrAlaGlnGlnGlyAsn 240
DB 661 TTAACAGATTACCACCAACAGGCGGATTTATCTACGCATTATGCCCAACAGGGAAC 720
QY 241 HisArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPheLys 260
DB 721 CACCGTTTATTTGAAGCCCAACCACTGTTTGGAGGGGTATCTTCCCTGAAATTTAA 780
QY 261 AspSerProArgTrpArgGlnThrGlyIleSerValLeuAsnThrGluIleLysLysGln 280

DB 781 GATTCAACCAAGATGGAGCCAAACCGGCATATCGGTGCTGAACACCCGAGATCAAAAAACAG 840
QY 281 ValTyrAlaAspGlyMetGlnPheGluLeuSerProIleTyrHisValAlaIleAsp 300
DB 841 GTTTATGCCGATGGGATGCAAGTTTGAACCTTTTCAACAAATTTACCATGTAGTCGATCGAT 900
QY 301 IlePheLeuLysAlaTyrGlySerAlaLysArgValAsnLeuGluLysGluPheProGln 320
DB 901 ATCTTCTTAAAGCCCTATGGTTCTGCAAAACAGATTACCTTGAAAAAGAAATTTCCGCAA 960
QY 321 SerTyrValGlnThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuProAsp 340
DB 961 TCTTATGTACAAACTGTAGAAAATATGATTATGGCGCTCATCAGTATTTTCACTGCCAGAT 1020
QY 341 TyrAsnThrProMetPheGlyAspSerTrpIleThrAspLysAsnPheArgMetAlaGln 360
DB 1021 TATAACACCCCTATGTTTGGAGATTATGAGATTATGAGATATAAAATTTTCAAGATGGCACAG 1080
QY 361 PheAlaSerTrpAlaArgValPheProAlaAsnGlnAlaIleLysTyrPheAlaThrAsp 380
DB 1081 TTTGCCAGCTGGCCCGGCTTTTCCGCGCAACACCGCCATAAATATTTTGTCTACAGAT 1140
QY 381 GlyLysGlnGlyLysAlaProAsnPheLeuSerLysAlaLeuSerAsnAlaGlyPheTyr 400
DB 1141 GGCAAACAGGTAAAGCGCCTTAACCTTTTATCCAAAGCATTGAGCAATCGAGCTTTTAT 1200
QY 401 ThrPheArgSerGlyTrpAspLysAsnAlaThrValMetValLeuLysAlaSerProPro 420
DB 1201 AGTTTAGAAGCGGATGGGATAAATGCAACCGTTATGGTATTAAGGCCAGTCTCTCC 1260
QY 421 GlyGluPheHisAlaGlnProAspAsnGlyThrPheGluLeuPheIleLysGlyArgAsn 440
DB 1261 GGGGAAATTCATGCCCGCCGATTAACCGGACTTTTGAACCTTTTATAAAGGCCAGAAC 1320
QY 441 PheThrProAspAlaGlyValPheValTyrSerGlyAspGluAlaIleMetLysLeuArg 460
DB 1321 TTTACCCACAGACCGCGGGTATTTGTGTATAGCGCGACGAAGCCATCATGAACTCGCG 1380
QY 461 AsnTrpTyrArgGlnThrArgIleHisSerThrLeuThrLeuAspAsnGlnAsnMetVal 480
DB 1381 AACTGGTACCGTCAACCCGCATACACAGCAGCTTACACTGCACAAATCAAAATATGGTC 1440
QY 481 IleThrLysAlaArgGlnAsnLysTrpGluThrGlyAsnAsnLeuAspValLeuThrTyr 500
DB 1441 ATTACAAAGCCCGCGCAAAACAATGGGAAACAGGAAATAACCTTGATGTGCTTACCTAT 1500
QY 501 ThrAsnProSerTyrProAsnLeuAspHisGlnArgSerValLeuPheIleAsnLysLys 520
DB 1501 ACCAAACCAAGCTATCCGAATCTGGACCATCGAGCAGTGTACTTTTTCATCAACAAAAA 1560
QY 521 TyrPheLeuValIleAspArgAlaIleGlyGluAlaThrGlyAsnLeuGlyValHisTyr 540
DB 1561 TACTTCTGGTTCATCGATAGGGCAATAGGCGAAGCTTACCGAAACCTGGCGGCTACACTGG 1620
QY 541 GlnLeuLysGluAspSerAsnProValPheAspLysThrLysAsnArgValTyrThrThr 560
DB 1621 CAGCTTAAAGAGACAGCAACCCCTGTTTTCGATAGACAAAGAACCGGGTTTACACCACT 1680
QY 561 TyrArgAspGlyAsnAsnLeuMetIleGlnSerLeuAsnAlaAspArgThrSerLeuAsn 580
DB 1681 TACAGAGATGGTAAACAACCTGATGATCCAACTCGTTGAATCGGACAGGACCGCTCAAT 1740
QY 581 GluGluGluGlyLysValSerTyrValTyrAsnLysGluLeuLysArgProAlaPheVal 600
DB 1741 GAAGAAAGAGGAAAGGATCTTATGTTTACAATAGAGAGCTGAAAGAGCTGCTTTCGTA 1800
QY 601 PheGluLysProLysLysAsnAlaGlyThrGlnAsnPheValSerIleValTyrProTyr 620
DB 1801 TTTGAAAGACCTTAAAGAAAGATCCCGCACACAAAAATTTTGTGAGTATAGTTATCCATAC 1860
QY 621 AspGlyGlnLysAlaProGluIleSerIleArgGluAsnLysGlyAsnAspPheGluLys 640

Db 1861 GACGGCCAGAGGCTCCAGAGATCAGCATACGGGAAACAAAGGCAATGATTTTGAGAAA 1920
QY 641 GlyLysLeuAsnLeuThrLeuThrIleAsnGlyLysGlnGlnLeuValLeuValPro 659
Db 1921 GGCAGCTTAATCTACCCCTTACCATTAACGAAACAAACAGCTTGTTGGTTCT 1977

RESULT 2
US-08-900-951-3
Sequence 3, Application US/08900951
Patent No. 5919693
GENERAL INFORMATION:
APPLICANT: Su, Hongsheng
APPLICANT: Blain, Francoise
APPLICANT: Bennett, Clark
APPLICANT: Gu, Kangfu
APPLICANT: Zimmermann, Joseph
APPLICANT: Musil, Roy
TITLE OF INVENTION: Nucleic Acid Sequences And Expression
TITLE OF INVENTION: Systems For Heparinase II And Heparinase III Derived From
TITLE OF INVENTION: Flavobacterium heparinum
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hale and Dorr
STREET: 1455 Pennsylvania Avenue, N.W.
CITY: Washington, D.C.
COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC DOS/MS DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/900,951
FILING DATE:
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/258,639
FILING DATE: 10 JUNE 1994
ATTORNEY/AGENT INFORMATION:
NAME: Healey, William J.
REGISTRATION NUMBER: 36,160
REFERENCE/DOCKET NUMBER: 104385.116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)942 8400
TELEFAX: (202)942 8484
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1980 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-900-951-3

Alignment Scores:
Pred. No.: 0 Length: 1980
Score: 3494.00 Matches: 659
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 2 Gaps: 0

US-09-802-285A-2 (1-659) x US-08-900-951-3 (1-1980)

QY 1 MetThrThyLysIlePheLysArgIleIleValPheAlaValIleAlaLeuSerSerGly 20
Db 1 ATGACTACGAAATTTTAAAGGATCATGTATTGCTGTAATTGCCCTATCGTCGGA 60
QY 21 AsnIleLeuAlaGlnSerSerIleThrArgLysAspPheAspHisIleAsnLeuGlu 40
Db 61 AATATACTTGCAAAAGCTCTCCATTAACAGGAAAGATTTTGACCACATCAACCTTGAG 120

QY 41 TyrSerGlyLeuGluLysValAsnLysAlaValAlaAlaGlyAsnTyrAspAspAlaAla 60
Db 121 TATTCGGACTGGAAGAGTTAATAAGCAGTTTCTCCGCGCACTATGACGATGCGGCC 180
QY 61 LysAlaLeuLeuAlaTyrTyrArgGluLysSerLysAlaArgGluProAspPheSerAsn 80
Db 181 AAAGCATTTACTGGCATACTACAGGAAAGAAAGTAAGCCAGGGAACCTGATTTTCAGTAAT 240
QY 81 AlaGluLysProAlaAspIleArgGlnProIleAspLysValThrArgGluMetAlaAsp 100
Db 241 GCAGAAAGCCTCGCATATACGCCAGCCCATAGATGAAGGTTACGGGTGAATGGCCGAC 300
QY 101 LysAlaLeuValHisGlnPheGlnProHisLysGlyTyrGlyTyrPheAspTyrGlyLys 120
Db 301 AAGGCTTTGGTCCACCAGTTTCAACCCGCAAAAGGCTACGGCTATTTTGATTTATGGTAAA 360
QY 121 AspIleAsnTrpGlnMetTrpProValLysAspAsnGluValArgTrpGlnLeuHisArg 140
Db 361 GACATCAACTGGCAGATGTGGCCGTAAAGACAAATGAAGTACGCTGGCAGTTGCACCGT 420
QY 141 ValLysTrpTrpGlnAlaMetAlaLeuValTyrHisAlaThrGlyAspGluLysTyrAla 160
Db 421 GTAAATATGGTCAGGCTATGGCCCTGGTTTATCACGCTACGGGCGCATGAAAAATATGCA 480
QY 161 ArgGluTrpValTyrGlnTyrSerAspTrpAlaArgLysAsnProLeuGlyLeuSerGln 180
Db 481 AGAGATGGGTATATCAGTACAGCATTTGGCCGAGAAACCCATTTGGCCCTGTGCGAG 540
QY 181 AspAsnAspLysPheValTrpArgProLeuGluValSerAspArgValGlnSerLeuPro 200
Db 541 GATAATGATAAAATTTGTGTGGCGGCCCTTGAAGTGTGCGACAGGGTACAAAGTCTTCCC 600
QY 201 ProThrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGluPhe 220
Db 601 CCAACCTTCAGCTTATTTGTAACTCCGAGCCCTTTACCCAGCCCTTTTAAATGGAATTT 660
QY 221 LeuAsnSerTyrHisGlnGlnAlaAspTyrLeuSerThrHisTyrAlaGluGlnGlyAsn 240
Db 661 TTAAACAGTTACCAACAGCCGCGATTATTAATCTACGCATTATGCCGAACAGGAAAC 720
QY 241 HisArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPheLys 260
Db 721 CACCGTTTATTTGAAGCCCAACGCAACTGTTTGCGAGGGTATCTTTCCCTGATTTAA 780
QY 261 AspSerProArgTrpArgGlnThrGlyLysSerValLeuAsnThrGluLysLysGln 280
Db 781 GATTCCACCAAGATGGAGGCAACCCGCATATCGTGTGTAACACCGAGATCAAAAAACAG 840
QY 281 ValTyrAlaAspGlyMetGlnPheGluLeuSerProIleTyrHisValAlaAlaIleAsp 300
Db 841 GTTTATGCCGATGGGATGAGTTTGACCTTTTACCCTTTTACCATGTAGTGCCTCGAT 900
QY 301 IlePheLeuLysAlaTyrGlySerAlaLysArgValAsnLeuGluLysGluPheProGln 320
Db 901 ATCTTCTTAAAGGCCCTATGGTTCTGCAAAACAGTTTAACCTTGAAAAAGAAATTTCCGCAA 960
QY 321 SerTyrValGlnThrValGluAsnMetIleMetAlaLeuLysSerIleSerLeuProAsp 340
Db 961 TCTTATGTACAAACTGTAGAAAATATGATTATGGCGCTGATGATGATTTTCACTGCCAGAT 1020
QY 341 TyrAsnThrProMetPheGlyAspSerTrpIleThrAspLysAsnPheArgMetAlaGln 360
Db 1021 TATACACCCCTATGTTTGGAGATTTCATGATTTACAGATAAAATTTTCAGGATGGCAG 1080
QY 361 PheAlaSerTrpAlaArgValPheProAlaAsnGlnAlaIleLysTyrPheAlaThrAsp 380
Db 1081 TTTTGGCAGCTGGCGCGGTTTTCGCGCAACACCGCCATTAATATTTTGTCTACAGAT 1140
QY 381 GlyLysGlnGlyAlaProAsnPheLeuSerLysAlaLeuSerAsnAlaGlyPheTyr 400
Db 1141 GGCACAAACAGGTAAGCGCGCTAACTTTTATCCAAAGCATTTGAGCAATTCAGGCTTTTAT 1200
QY 401 ThrPheArgSerGlyTrpAspLysAsnAlaThrValMetValLeuLysAlaSerProPro 420

Db 1201 AGTTTAGAGCGGATGGATATAAATGCAACCGTTATGTTATTAAGCCAGTCTCTCC 1260
QY 421 GlyGluPheHisAlaGlnProAspAsnGlyThrPheGluLeuPheIleLysGlyArgAsn 440
Db 1261 GGGGAATTTCTGCGCCAGCGGATAAACGGGACTTTTGAACCTTTTATATAAGGCGAGAAAC 1320
QY 441 PheThrProAspAlaGlyValPheValTyrSerGlyAspGluAlaIleMetLysLeuArg 460
Db 1321 TTATCCCGCAGACCGCGGGTATTTGTTATACGGCGACGAGCCATCATGAACCTGGCG 1380
QY 461 AsnTrpTyrArgGlnThrArgIleHisSerThrLeuThrLeuAspAsnGlnAsnMetVal 480
Db 1381 AACTGGTACCGTCAACCCGCGATACACAGCAGCTTACACTGCACATCAATAATGGTTC 1440
QY 481 IleThrLysAlaArgGlnAsnLysTrpGluThrGlyAsnAsnLeuAspValLeuThrTyr 500
Db 1441 ATTACCAAGCCCGCAACCAATGGGAACAGGAAATAACCTTGTATGCTTACCTAT 1500
QY 501 ThrAsnProSerTyrProAsnLeuAspHisGlnArgSerValLeuPheIleAsnLysLys 520
Db 1501 ACCAACCCAGCTATCCGAACTTGGACCATCAGCGAGTGTACTTTTTCATCAACAAAAA 1560
QY 521 TyrPheLeuValIleAspArgAlaIleGlyGluAlaThrGlyAsnLeuGlyValHisTrp 540
Db 1561 TACTTTCTGTCATCGATAGGCAATAGCGAGCTACCGGAACCTGGCGGTACACTGG 1620
QY 541 GlnLeuLysGluAspSerAsnProValPheAspLysThrLysAsnArgValTyrThr 560
Db 1621 CAGCTTAAAGAGACAGCAACCTGTTTTCGATAGACAAAGAACCGGGTTTACACCACT 1680
QY 561 TyrArgAspGlyAsnAsnLeuMetIleGlnSerLeuAsnAlaAspArgThrSerLeuAsn 580
Db 1681 TACAGAGATGGTAACCACTGATGATCCCAATCGTTGAATCGGACAGGACCACTCAAT 1740
QY 581 GluGluGluGlyLysValSerTyrValTyrAsnLysGluLeuLysArgProAlaPheVal 600
Db 1741 GAAGAAGAGGAAGAGTACTTATGTTTACATAAGGAGCTGAAGAAGACCTGCTTTCGTA 1800
QY 601 PheGluLysProLysLysAsnAlaGlyThrGlnAsnPheValSerIleValTyrProTyr 620
Db 1801 TTTGAAAGCCCTAAAGAAAGTACCGGCACACAAAATTTTGTCAATATGTTTATCCATAC 1860
QY 621 AspGlyGlnLysAlaProGluIleSerIleArgGluAsnLysGlyAsnAspPheGluLys 640
Db 1861 GACGGCCAGAGCTCCAGAGATCAGCATACGGGAAACAGGGCAATGATTTTGAGAAA 1920
QY 641 GlyLysLeuAsnLeuThrLeuThrIleAsnGlyLysGlnGlnLeuValLeuValPro 659
Db 1921 GCGAAGCTTAATCTAACCTTACCATTAACGGAAACAAACAGCTTGTGTGTTCTCT 1977

RESULT 3

PCT-US95-07391A-3
; Sequence 3, Application PC/TUS9507391A
; GENERAL INFORMATION:
; APPLICANT: IBEX TECHNOLOGIES and
; APPLICANT: ZIMMERMANN, Joseph
; TITLE OF INVENTION: Nucleic Acid Sequences And Expression
; TITLE OF INVENTION: Systems For Heparinase II And Heparinase III Derived From
; TITLE OF INVENTION: Flavobacterium heparinum
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hale and Dorr
; STREET: 1455 Pennsylvania Avenue, N.W.
; CITY: Washington, D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US95/07391A
; FILING DATE: 09-JUNE-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/258,639
; FILING DATE: 10 JUNE 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Hollie L.
; REGISTRATION NUMBER: 31,321
; REFERENCE/DOCKET NUMBER: 104385.116PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)942-8400
; TELEFAX: (202)942-8484
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1980 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
PCT-US95-07391A-3

Alignment Scores:

Pred. No.: 0 Length: 1980
Score: 3494.00 Matches: 659
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 5 Gaps: 0

US-09-802-285A-2 (1-659) x PCT-US95-07391A-3 (1-1980)

QY 1 MetThrThrLysIlePheLysArgIleIleValPheAlaValIleAlaLeuSerSerGly 20
Db 1 ATGACTACGAAATTTTAAAGAGGATCATGTTATTTGCTGTAATTCCTATCGTCGGGA 60
QY 21 AsnIleLeuAlaGlnSerSerSerIleThrArgLysAspPheAspHisIleAsnLeuGlu 40
Db 61 NATATCTTGCAACAAGCTCTTCATTACCAGGAAGATTTTGACCACATCAACCTTGAG 120
QY 41 TyrSerGlyLeuGluLysValAsnLysAlaValAlaAlaGlyAsnTyrAspAlaAla 60
Db 121 TATTCGGACTGGAAAAAGGTTAATAAGCAGTTGCTGCGGCAACTATCAGCATGGGCC 180
QY 61 LysAlaLeuLeuAlaTyrThrArgGluLysSerLysAlaArgGluProAspPheSerAsn 80
Db 181 AAAGCATTACTGGCATACTACAGGAAAAAGTAAGGCCAGGAACTGATTTCAGTAAT 240
QY 81 AlaGluLysProAlaAspIleArgGlnProIleAspLysValThrArgGluMetAlaAsp 100
Db 241 GCAGAAAAGCCTGCCGATATACGCCAGCCCATAGATAAGTTACGCGTGAATGGCCGAC 300
QY 101 LysAlaLeuValHisGlnPheGlnProHisLysGlyTyrGlyTyrPheAspTyrGlyLys 120
Db 301 AAGGCTTTGGTCCACCGAGTTTCAACCGCAACAAGGCTACGGCTATTTTATGATTGTAA 360
QY 121 AspIleAsnTrpGlnMetTrpProValLysAspAsnGluValArgTrpGlnLeuHisArg 140
Db 361 GACATCACTGGCAGATGTGGCGGTAAAGACATCAAGTACGCTGGCAGTTGCCCGT 420
QY 141 ValLysTrpTrpGlnAlaMetAlaLeuValTyrHisAlaThrGlyAspGluLysTyrAla 160
Db 421 GTAAAAATGGTGGCAGGCTATGCGCTGTTTATCAGCTACGGGCGATGAAAAATATGCA 480
QY 161 ArgGluTrpValTyrGlnTyrSerAspTrpAlaArgLysAsnProLeuGlyLeuSerGln 180
Db 481 AGAGAATGGGTATATCAGTACAGCGATTGGGCCAGAAAAAACCATTGGGCTGTGCGAG 540
QY 181 AspAsnAspLysPheValTrpArgProLeuGluValSerAspArgValGlnSerLeuPro 200
Db 541 GATATGATGAATTTGTGTGGCGGCCCTTGAAGTGTGGACAGGAGGTCAACAGCTCTCCC 600
QY 201 ProThrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGluPhe 220

Db 601 CCAACCTTCAGCTATTATTGTAACCTCCGACGCTTTACCCAGCCTTTTAAAGCAATTT 660
QY 221 LeuAsnSerTyrHisGlnGlnAlaAspTyrLeuSerThrHisTyrAlaGluGlnGlyAsn 240
Db 661 TTAACAGATTACCACCAACAGGCGGATATTATTATCTACGCATTATGCGGAACAGGGAAC 720
QY 241 HisArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPheLys 260
Db 721 CACGTTTATTGAGCCCAACGCACTTGTGTCAGGGGTATCTTCCCTGGAATTTAAA 780
QY 261 AspSerProArgTrpArgGlnThrGlyIleSerValLeuAsnThrGluIleLysLysGln 280
Db 781 GATTCCCAAGATGGAGCAACCGGCATATCGGTGCTGAACCCGAGATCAAAAACAG 840
QY 281 ValTyrAlaAspGlyMetGlnPheGluLeuSerProIleTyrHisValAlaIleAsp 300
Db 841 GTTTATGCCGATGGGATGCAGTTTGAACTTTCACCAATTTACCAATTTAGCTGCGCATCGAT 900
QY 301 IlePheLeuLysAlaTyrGlySerAlaLysArgValAsnLeuGluLysGluPheProGln 320
Db 901 ATCTTCTTAAGCCCTATGTTCTGCAACAGAGTTAACCTTGAAAAGAAATTTCCGCAA 960
QY 321 SerTyrValGlnThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuProAsp 340
Db 961 TCTTATGTACAACTGTAGAAATATGATTATGCGCTGATCAGTATTTCACGCGCAGAT 1020
QY 341 TyrAsnThrProMetPheGlyAspSerTrpIleThrAspLysAsnPheArgMetAlaGln 360
Db 1021 TATAACACCCCTATGTTTGGAGATTCATGATTACAGATAAAATTTTCAGGATGGCAG 1080
QY 361 PheAlaSerTrpAlaArgValPheProAlaAsnGlnAlaIleLysTyrPheAlaThrAsp 380
Db 1081 TTTGCCAGCTGGCCCGGGTTTCCCGCAACAGGCCCAATAAATTTTGTCTACAGAT 1140
QY 381 GlyLysGlnGlyLysAlaProAsnPheLeuSerLysAlaLeuSerAsnAlaGlyPheTyr 400
Db 1141 GGCAACAAGGTAAAGCGCTAACTTTTATCCAAAGCATTTAGCAATGCGAGCTTTTAT 1200
QY 401 ThrPheArgSerGlyTrpAspLysAsnAlaThrValMetValLeuLysAlaSerProPro 420
Db 1201 ACCTTTAGAGCGGATGGGATAAAATGCAACCGTTTATGTTATTAAGCGCAGTCCTCC 1260
QY 421 GlyGluPheHisAlaGlnProAsnAsnGlyThrPheGluLeuPheIleLysGlyArgAsn 440
Db 1261 GGGGAATTCATGCCACCGCGGATTAACGGACITTTTGAACCTTTTATAAAGGCGAGAAC 1320
QY 441 PheThrProAspAlaGlyValPheValTyrSerGlyAspGluAlaIleMetLysLeuArg 460
Db 1321 TTTACCCACAGACCGCGGGTATTGTTGTATAGCGGCGAGCAAGCCATCATGAACCTCGG 1380
QY 461 AsnTrpTyrArgGlnThrArgIleHisSerThrLeuThrLeuAspAsnGlnAsnMetVal 480
Db 1381 AACTGTGTACGTCAAAACCCCATACACAGCAGCGTTTACCTCGACAAATCAAAATATGTC 1440
QY 481 IleThrLysAlaArgGlnAsnLysTrpGluThrGlyAsnAsnLeuAspValLeuThrTyr 500
Db 1441 ATTACCAAGCCCGCAAAACAATGGAAACAGGAATAACCTTGATGTGCTTACCTAT 1500
QY 501 ThrAsnProSerTyrProAsnLeuAspHisGlnArgSerValLeuPheIleAsnLysLys 520
Db 1501 ACCAAACCAAGTATCGAATCTGGACCATCAGCGCAGTGTACTTTTCATCAACAAAAA 1560
QY 521 TyrPheLeuValIleAspArgAlaIleGlyGluAlaThrGlyAsnLeuGlyValHisTrp 540
Db 1561 TACTTTCTGTGTATCGATAGGCAATAGGCGAAAGCTACCGGAAACCTGGCGGTACACTGG 1620
QY 541 GlnLeuLysGluAspSerAsnProValPheAspLysThrLysAsnArgValTyrThrThr 560
Db 1621 CAGCTTAAAGAGACAGCAACCTGTTTTCGATTAAGCAAAAGAACCGGGTTTACACCACT 1680
QY 561 TyrArgAspGlyAsnAsnLeuMetIleGlnSerLeuAsnAlaAspArgThrSerLeuAsn 580

Db 1681 TACAGAGATGTTAACACCTGATGATCCCAATCGTTGAATCGGACAGACAGCCTCAAT 1740
QY 581 GluGluGluGlyLysValSerTyrValTyrAsnLysGluLeuLysArgProAlaPheVal 600
Db 1741 GAAGAAGAAGAAAGGTATCTTATGTTTACAATAAGAGAGCTGAAAAGACCTGCTTTCGTA 1800
QY 601 PheGluLysProLysLysAsnAlaGlyThrGlnAsnPheValSerIleValTyrProTyr 620
Db 1801 TTTGAAAAGCCTAAAAGAAATGCCGACACAAAATTTTGTAGTATATGTTATCCATAC 1860
QY 621 AspGlyGlnLysAlaProGluIleSerIleArgGluAsnLysGlyAsnAspPheGluLys 640
Db 1861 GACGCCAGAGGCTCCAGAGATCAGATACGGAACAAAGGCAATGATTTTGAGAA 1920
QY 641 GlyLysLeuAsnLeuThrLeuThrIleAsnGlyLysGlnLeuValLeuValPro 659
Db 1921 GGCAAGCTTAATCAACCTTACCATTAACGGAACAAACAGCTTGTGTTGTTCT 1977

RESULT 4

US-08-961-527-186/c
; Sequence 186, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 186:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3763 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-527-186

Alignment Scores:
Pred. No.: 5,63e-10 Length: 3763
Score: 176.00 Matches: 116
Percent Similarity: 34.63% Conservative: 80
Best Local Similarity: 20.49% Mismatches: 231
Query Match: 5.04% Indels: 140
DB: 4 Gaps: 20

US-09-802-285A-2 (1-659) x US-08-961-527-186 (1-3763)

QY 71 SerLysAlaArgGluProAspPheSerAsnAlaGluLysProAlaAspIleArgGlnPro 90

Db 2173 TCMAAGACTATCAGAAAGTCAAGAGTTCCTTGAACGCTTGATGACCAATCGTTTATG 2114

QY 91 IleAspLysValThrArgGluMetAlaAspLysAlaLeuValHisGlnPheGlnProHis 110
 Db 2113 TTGTAGT---CCTTGGATAGGAGCCTTGTTCAAATCCATCAATCCAGCG--- 2060
 QY 111 LysGlyTyrGlyTyrPheAspTyrGlyLysAspIleAsnTrpGlnMetTrpProValLys 130
 Db 2059 -----ATGGTATGGGATCAAGTAATTTGAAGAT 2033
 QY 131 AspAsnGluValArgTrpGlnLeuHisArgValLysTrpGlnAlaMetAlaLeuVal 150
 Db 2032 GATCCAGAATGCTTATATGCTCAATCGACAGAATATCTCTGCGATTTATGATAGG 1973
 QY 151 TyrHisAlaThrGlyAspGluLysTyrAlaArg-----GluTrp 163
 Db 1972 TATCTGGTAGAGGATGAGACTATATCAAAAGTCAAGTCTCTTCTATTTGATGG 1913
 QY 164 ValTyrGlnTyrSerAspTrpAlaArgLysAsnProLeuGlyLeuSerGlnAspAsnAsp 183
 Db 1912 ATTGACGAGTGAGAGAAATTTCTCTCAATCTTGATGACTAGAACCTTGATACGGGT 1853
 QY 184 -----LysPheValTrpArgProLeuGluValSerAspArgValGlnSerLeuProPro 201
 Db 1852 ATCGTCTCTTACTTGGTTGAAACTA----- 1826
 QY 202 ThrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGluPheLeu 221
 Db 1825 ---CCTTCTCTCTCTTGAATTTGACTTGTAGAGGAGAAAGAACTAGAGAAATTTG 1769
 QY 222 AsnSerTyrHisGlnAlaAspTyrLeuSerThrHisTyrAlaGluGln----- 238
 Db 1768 GTCACTAGAAAAGCAGATTCACCTTATGAAAAGCTACTATCGCCCAAGTACACCTT 1709
 QY 239 GlyAsnHisArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGlu 258
 Db 1708 AGTAACCTGGGGATTTTCAACAATCCGATGCTTGTCTATCTATCTTCTTTCAGAT 1649
 QY 259 PheLysAspSerProArgTrpArgGlnThrGlyIleSerValLeuAsnThrGluLeuLys 278
 Db 1648 AAGATGGAICTAGAGAAGCTTACCATTTTGCTTCAGAGAGTTGAAACAGCAATTTAG 1589
 QY 279 LysGlnValTyrAlaAspGlyMetGlnPheGluLeuSerProIleTyrHisVal----- 296
 Db 1588 ACACAGATTTAGGAGATGGAAGCCAGTTTGAACAGTCGATCTCTATCATGTAGAGTT 1529
 QY 297 -----AlaAlaIleAspIlePheLeuLysAlaTyrGlySerAlaLysArgValAsnLeu 314
 Db 1528 TATAAGCCTTCTGATTTGTCTCTC-----TTG 1499
 QY 315 GluLysGluPheProGlnSerTyrValGlnThrValGluAsnMetIleMetAlaLeuIle 334
 Db 1498 CTTCCAGACTTCAAGATAGTTACCAAGAGTTCTGTGAAAGATGCGACCTATATCAA 1439
 QY 335 SerIleSerLeuProAspTyrAsnThrProMetPheGlyAspSerTrpIleThrAspLys 354
 Db 1438 ATGATGACAGGCTTAGATGAGCGACTTTGGCTTTTGGTGATAGCGATTTCTACAGAAAG 1379
 QY 355 AsnPheArgMetAlaGlnPheAla----- 362
 Db 1378 ACAGAAATTTGAGCCTCTCTCTGTGGTTTGAACCCAGAGACCTTCTTAACGGTCTG 1319
 QY 363 -----SerTrpAlaArgValPheProAlaAsnGlnAlaIleLysTyrPhe 377
 Db 1318 GATGTTAAAGTTGATTGTGCTTGTCTTCTTCTGGGGCGAGAAAGGTCAAGCGACTG 1259
 QY 378 AlaThrAspGlyLysGlnGlyLysAlaProAsnPheLeuSerLysAlaLeuSerAla 397
 Db 1258 CAGGAATTTGAAAAGAGAGCTTGGCGACCTAAG-----TCCATGATCTTTGAAGACTCT 1205
 QY 398 GlyPheTyrThrPheArgSerGlyTrpAspLysAsnAlaThrValMetValLeuLysAla 417
 Db 1204 GGACATGCTCGCATTAAGGAT-----GAACATCGTTATCTATTTTCAAAAT 1157

QY 418 SerProProGlyCluPheHisAlaGlnProAspAsnGlyThrPheGluLeuPheIleLys 437
 Db 1156 GTCCTCGTAGGAGTCCCATAGTCATAGTAGAGAGAAATAGTTTTGCTTACATATCAA 1097
 QY 438 GlyArgAsnPheThrProAspAlaGlyValPheValTyrSerGlyAspGluAlaIleMet 457
 Db 1096 GGCCCAACCTATTTCATAGATCTGGCGCTTATCTTATCG--GGAGATATATGAACGTTA 1038
 QY 458 LysLeuArgAsnTrpTyrArgGlnThrArgIleHisSerThrLeuThrLeuAspAsnGln 477
 Db 1037 TCTCTAAAGAGTGTGGAGTCATTCGACCTGCAATGT----- 999
 QY 478 AsnMetValIleThrLysAlaArgGlnAsnLysTrpGlu---ThrGlyAsnAsnLeuAsp 496
 Db 998 -----AGATGGAAAGCTCCGGAAGAAATCACGGG 969
 QY 497 ValLeu-ThrTyr----- 500
 Db 968 ATCCTGGGAATATGAATACTATCTCTCCTCCTGTTTGTCCACATAAAGAGGAGGG 909
 QY 501 -----ThrAsnProSerTyrProAsnLeuAspHi 510
 Db 908 ATGCAATATATGAGGGGCTTATGTGCAGCAACCTGATTTGCCCTTATCTT---CA 852
 QY 510 sGlnArgSerValLeuPheIleAsnLysLysTyrPheLeuValIleAsp-----ArgAl 528
 Db 851 CAGGAGAAAATCCTCATGTTGGTAGAGGATCTCGGCTCTTGGTAGATCATCAGGTG 792
 QY 528 alleGlyGluAlaThrGlyAsnLeuGlyValHisTrpGlnLeuLysGluAspSerAsnPr 548
 Db 791 TCAAGGTCAA-----CATGAGGTGTGCACTCACTTT----- 761
 QY 548 oValPheAspLysThrLysAsnArgValTyrThrTyrArgAspGly-----AsnAs 566
 Db 760 -ATCCTTGACAGGAT-----GTGACCTATCAAGATGGGAAAATCAATCA 717
 QY 566 nLeuMetIleGlnSer 571
 Db 716 GTTGAGACTATGGAGT 701
 RESULT 5
 US-09-134-001C-795
 ; Sequence 795, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; PRIOR FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 795
 ; LENGTH: 2097
 ; TYPE: DNA
 ; ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-795
 Alignment Scores:
 Pred. No.: 2,66e-05 Length: 2097
 Score: 132.50 Matches: 140
 Percent Similarity: 34.59% Conservative: 108
 Best Local Similarity: 19.53% Mismatches: 260
 Query Match: 3.79% Indels: 210
 DB: 4 Gaps: 30
 US-09-802-285A-2 (1-659) x US-09-134-001C-795 (1-2097)
 QY 9 IleIleValPheAlaValIleAlaLeuSerSerGlyAsnIle---LeuAlaGlnSerSer 27

```

Db 94 GTATTATCTTGTCTAGTCGTATTGAGATTAGTTATTACAAATAGCACAGGATCT 153
QY 28 SerIleThrArg-----LysAspPheAspHisIleAsnLeuGluTyrSerGlyLeuGlu 45
Db 154 CATTACAACAATTAATCAAAAGCATGAGAAACATAACT----- 192
QY 46 LysValAsnLysAlaValAlaAlaGlyAsn-----Tyr 56
Db 193 ---GTTAATGATCAGTACCAGAGCGCGAATACTAGATGAATGCGCAAGTACTAGTT 249
QY 57 AspAspAlaLysAlaLeuLeuAlaTyrTyrArgGluLysSerLysAlaArgGluPro 76
Db 250 GATAATGCTTCAAGATGCTATTACATACACTAGAAACCGTAAACATCACAAAGGAA 309
QY 77 AspPheSerAsnAlaGluLysProAlaAsp---IleArgGlnProIleAspLysValThr 95
Db 310 ATGTTAAATATCTGCTAGAAACTGACAGATTAAATTAATTAATGATGATACAGATAAAATTA 369
QY 96 ArgGluMetAlaLysAlaLeuValHisGlnPheGlnProHisLysGlyTyrGlyTyr 115
Db 370 GAGAGA---GATAAA----- 381
QY 116 PheAspTyrGlyLysAspIleAsnTyrGlnMetTyrPro-----ValLysAspAsn 132
Db 382 -----AAGGATTTTGGATTCAATGATGATCCGTCATTTCGCTAAAGATT-AAT 428
QY 133 GluValArgTyrGlnLeuHisArgValLysTyrTyrGln----- 145
Db 429 GAGAAAAGA-----ACAATTAATGTTAGAGGATGGCAGTATTTCACAAGACCAATTTGA 482
QY 146 ----- 147
Db 483 TACCCCACTTAGAGATAAATAGAGAAAACAAATTAACAGTTTAACATAAAAGATT 542
QY 148 AlaLeuValTyrHisAlaThrGlyAspGluLysTyrAla----- 160
Db 543 GCAAGTTTATGCAATTTATCGGAAATGAACGTCGGTCAACTAGATCCTCAACAAAT 602
QY 161 -----ArgGlu----- 162
Db 603 TAAAAATGAAGACGTAAAGCGAGAAAGAAATATGACGCGTATCAACAGAGCTTTCTAAATT 662
QY 163 ---TipValTyrGlnTyrSer-AspTyrAlaArgLysAsnProLeuGlyLeuSerGlnAs 181
Db 663 ACCTGGTGTAATACTACAAATGATGGGTAGGATAGAAATACCCATACCGT-----GA 713
QY 181 pAsnAspLysPheValTyrArgProLeuGluValSerAspArgValGlnSerLeuProPr 201
Db 714 TACTTTAAGAGGCATATTTTGA-----GATGTGTCGACTTCGACTGAAGGTATACCTAA 767
QY 201 oThrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGluPheLe 221
Db 768 AGAATTAAC-----GAACAATATTT 788
QY 221 uAsnSerTyrHisGlnGlnAlaAspTyrLeuSerThrHisTyrAlaGluGlnGlyAsnHi 241
Db 789 ATCAAAAGTTTATTCAGAAATGATCGGTCGGTAAATCTTACTTGAATATCAATACGA 848
QY 241 sArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPheLysAs 261
Db 849 AGATGTTCTTAAAGGCACGAGAAACAAATG-----AAATATACAAC 890
QY 261 pSerProArgTyrArgGlnThrGlyIleSerValLeuAsnThrGluLysLysGlnVa 281
Db 891 TGATAAATCTGGAAGAGTAATAAGTTCAGAAGTACTTAATCTCGCTGCTCAAGA----- 942
QY 281 lTyrAlaAspGlyMetGlnPheGluLeuSerProIleTyrHisValAlaAlaIleAspIl 301
Db 943 -----GGTCATGATTACAAATTAAC-----ATAGATAT 971
QY 301 ePheLeuLysAlaTyrGlySerAlaLysArgValAsn-----LeuGluLysGluPhePr 319
| | | | |

```

```

Db 972 TGATTTCAG-----AAAAAGTAGAATCTTTATTAGAAAAACAATTTTC 1016
QY 319 oGlnSerTyrValGlnThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuPr 339
Db 1017 TAAATTACGTAGTCAAGGTCTAAGGATATGGACAATCGCTTAATGTTGTCTCCAAAATCC 1076
QY 339 oAspTyrAsnThrProMetPheGlyAspSerTyrIleThrAspLysAsnPheArgMetAl 359
Db 1077 TAAAAATGGAGACATTTCTCGCTATTGCGAGGAAGCAAAATGTAGAACAGGTAAACCAA 1136
QY 359 aGlnPhe-----AlaSerTyrAlaArgValPheProAlaAsnGlnAlaIleLys---Ty 376
Db 1137 AGATTATGATATCGGCACTTTACAGCTCAATACACAGTAGGTCTTCACGTAAAGCAGG 1196
QY 376 rPheAlaThrAspGlyLysGlnGlyLysAlaProAsn-----PheLeuSerLys 392
Db 1197 AACATTATTAGCTGGATACCAAAATAAAGCTATTAAATGTTGGAGAACTACTGTTAGATGA 1256
QY 392 sAlaLeuSerAsnAlaGlyPheTyrThrPheArgSerGlyTyrAspLysAsnAla----- 410
Db 1257 GCCATTAAAAATCCAGGTGGTTTAACTAAGCGTCTTATTATTAATAAAAAATGGTCAATG 1316
QY 411 -----ThrVa 412
Db 1317 ATCTATCGATATAAACAACACTTATGCATTTCATCAACGTATATACATGTTTAAACCGC 1376
QY 412 lMetValLeuLysAlaSerProGlyGluPheHisAlaGlnProAspAsnGlyThrPh 432
Db 1377 ACTTAAATTAGCAGGTGACCCCTATCTTCCAGGTATGTCATTACCTAATAAT----- 1428
QY 432 eGluLeuPheIleLysGlyArgAsnPheThrProAspAlaGlyValPheValTyrSerGl 452
Db 1429 -----ATAGCAGATGCTGGT----- 1443
QY 452 yAspGluAlaIleMetLysLeuArgAsnTyrTyrArgGlnThrArgIleHisSerThrLe 472
Db 1444 -----CGTAAATTCGTTAAAGGTTTAAATCAAGTAGGTCTTCGCTTAAAC 1490
QY 472 uThrLeuAspAsnGlnAsnMetValIleThrLysAlaArgGlnAsnLysTyrGluThrGl 492
Db 1491 AGGTATTGACTTACCGAACGAAACGCCAGCCCAATAGAACCCCTTAATAATAATCCTGG 1550
QY 492 yAsnAsnLeuAspValLeuThrTyrThrAsnProSerTyrProAsnLeuAspHisGlnAr 512
Db 1551 TAACTATTAGACTTACTGCTATTGCGCAATACGACATATACACACTTCAGTTGTCCTAT 1610
QY 512 gSerValLeuPheIleAsnLysLysTyrPheLeuVal-----IleAspArgAlaIl 529
Db 1611 ATACGTATCAACTATTGCTAATGATGCTATAGAAATCAACACATATTGGATTGCTAT 1670
QY 529 eGlyGluAlaThrGlyAsnLeuGlyValHisTyrGlnLeuLysGluAspSerAsnProVa 549
Db 1671 TTATGAATCTACTAAT-----AAAGATCAAACTGGTCCATT 1706
QY 549 lPheAspLysThrLysAsnArgValTyrThrTyrArgAspGlyValAsnLeuMetIl 569
Db 1707 AAAACCTAAATTAAGGTAATGTTTTTAATAAGGTAATAATACTCAATGACGAAATTA 1766
QY 569 eGlnSerLeuAsnAlaAspArgThrSerLeuAsnGluGluGlyLysValSerTyrV 589
Db 1767 AGAAGTTCAAGAGGCTTCAAAATGGCTTTCATTAAGAAAGCAAGGTACAGGTTATGCTAG 1826
QY 589 al-----TyrAsnLysGluLeuLysArgProAlaPheValPheGluLysP 604
Db 1827 TTTTGAATAATCTGTAGTACCTTCAGCTGCTAAACAGCAACTGCTGAAGTGTTCACAAGA 1886
QY 604 roLysLysAsnAlaGlyThrGlnAsnPheValSerIleValTyrPro 619
Db 1887 CGGAGAACCTAGAGTTTAACCTCAACAT---ATATCGGTTATGACCCCG 1930

```

RESULT 6

US-07-973-320-1

; Sequence 1, Application US/07973320

Patent No. 5286486
GENERAL INFORMATION:
APPLICANT: Payne, Jewel M.
TITLE OF INVENTION: No. 5286486el Bacillus thuringiensis Gene
TITLE OF INVENTION: Encoding a Coleopteran-Active Toxin
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM: disk
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/973,320
FILING DATE: 19921106
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/788,638
FILING DATE: 6-NOV-1991
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: MA68.C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3414 base pairs
TYPE: NUCLEIC ACID
STRAINEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus thuringiensis
STRAIN: dakota
INDIVIDUAL ISOLATE: HD511
IMMEDIATE SOURCE:
LIBRARY: Lamdagem (TM)-11 library of J.M. Fu
CLONE: 511
US-07-973-320-1

Alignment Scores:
Pred. No.: 0.00669 Length: 3414
Score: 123.50 Matches: 141
Percent Similarity: 35.92% Conservative: 95
Best Local Similarity: 21.46% Mismatches: 199
Query Match: 3.53% Indels: 224
DB: 1 Gaps: 39

US-09-802-285A-2 (1-659) x US-07-973-320-1 (1-3414)

QY 48 AsnLysAlaValAla-----AlaGlyAsnTyrAspAspAlaAlaLysAlaLeu 63
Db 361 AATAAGCACTTCGACAAATTAGAGGATTAGGAATAAC----- 399
QY 64 LeuAlaTyrArgGluLysSerLysAlaArgGluProAspPheSerAsnAlaGluLys 83
Db 400 TTAACAATATATCAACAG-----GCATTGAAGATTGGCTGAACAATCTGTATGAT 450
QY 84 ProAlaAspIleArgGlnProIleAspLysValThrArgGluMetAlaAspLysAlaLeu 103
Db 451 CCAGCACTATACACAGGTAGATCGTTT-----CGTATATAGAT---GCTTTA 501

QY 104 ValHisGlnPheGlnProHisLysGlyTyrGlyTyrPheAsp----- 117
Db 502 TTGGAATCATATATGCCGTCATTTAGGTTGCTGGATATGAATACCATTTACTAACAGTT 561
QY 118 TyrGlyLysAspIleAsnTrpGlnMet----- 126
Db 562 TAGGCACAAGCGCAACCTTCATCTAGCTTTTATTAAAGAGATTCTACTCTTTATGAGAT 621
QY 127 ---TrpProValLysAspAsnGluVal----- 134
Db 622 AAATGGGATTCACATCAGAACCAATTTAGGAAAAATTAATTCGTCAAAAGAAACATATC 681
QY 135 ---ArgTrpGlnLeuHisArgValLysTrpTyrGlnAla---MetAlaLeuValTyrHis 152
Db 682 TCTGAATATTTAAACCATTCGCTTAAAGTGTATATAGTGTCTTAGAGATTAACCGGT 741
QY 153 AlaThrGlyAspGlu-----LysTyrAlaArgGluTrpValTyrGlnTyr 167
Db 742 TCACCTTATGAACAATGGATAAATTAATAATCGTTTCGTAGAGAAATGATATTAATGGTA 801
QY 168 SerAspTrpAlaArgLysAsnProLeu-----GlyLeu 178
Db 802 TTAGATATTGCTGCTGTTATTTCTATTTATGACCTCGAATGTATTCAATGGAACAAGT 861
QY 179 SerGlnAspAsnAspLysPheValTrpArg-ProLeuGluValSerAspArgValGlnSe 198
Db 862 AGCAGTT-AAAGAGAGAGTGTATACCATTCATTTAGCTTGTCAATTTAGCAATCCAGA 920
QY 198 rLeuProThrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMe 218
Db 921 TATAGTCCAAAGTTTTCTCAGATGGAAAAATACTGCGTTTAAACAACACACACCTTGT-- 978
QY 218 tGluPheLeuAsnSerTyrHisGlnAlaAspTyrLeuSerThrHisTyrAlaGlu 238
Db 979 -----GATTATTAGATGAGCTTTATATATATAC 1007
QY 238 nGlyAsnHisArgLeuPhe-----GluAlaGlnArgAsnLeuPhe-----Al 252
Db 1008 ATCAAAATATAAAGCATTTTCACATGAGATTCAACACGACCTATTATTATGTTGTGTACA 1067
QY 252 aGlyValSerPheProGluPheLysAspSerProArgTrpArgGlnThrGlyIle-SerV 272
Db 1068 TAAGGTAGCTTTAAAAAATCGAGCAATCCCAATTTATAT---ACAACAGGCATATATGG 1124
QY 272 aLeuAsnThrGluIle-----LysLysGlnValTyrAlaAspGlyMet---GlnPheG 289
Db 1125 TAAACAAGTGGATATATTTTCATCAGAGCATATTCATTTAGAGGAATGATATCTATAG 1184
QY 289 luLeuSerProIleTyrHisValAlaAlaIleAspIlePheLeuLysAlaTyrGlySera 309
Db 1185 AACATTAGCAGCTCCATCAGTTGTAGTTTATCCGTATA---CTCAGAATTATGTTGTCCA 1241
QY 309 laLys----- 310
Db 1242 GCAAGTTGAGTTTACGGTTAAAGGGCATGTACATATTAGAGGAGATAACAAATATGA 1301
QY 311 ----ArgValAsnLeuGluLysGluPheProGlnSerTyrValGlnThrValGluAsnM 329
Db 1302 TCTGAGTATGATTCTATTGATCAATACCACGAGGAGAACCAATACACGAAAAATA 1361
QY 329 etIleMetAlaLeu-----IleSerIleSerLeuProAspTyrAsnThr- 343
Db 1362 CACTCATCGA-TTATGTGTCATGTACAGCTATATCTAAATCAACTCCGATTATGATAATG 1420
QY 344 -----ProMetPheGlyAspSerTrpIleThrAspLysAsnPheAspMetAlaGlnP 361
Db 1421 CTACTATCCGATCTTT-----TCTTGG-----ACGCATAGAGTCCGAGT 1462
QY 361 heAlaSerTrpAlaArgValPheProAlaAsn-----GlnAlaIleLysTyrP 377
Db 1463 AT-----TACATAGATCTATCCAAACAAATCAAAAAATTCACAGCTGTAAAAATGT 1516
QY 377 heAlaThrAspGlyLysGlnGly-----LysAlaProAsnPheLeuSerLysAlaLeuS 395

```
Db 1517 ATAACTAGTATCTATCTACAGTTGTCAAGGGCTGATTTACAGGTGGAGATTAG 1576
QY 395 eAsnAlaGlyPheTyrThrPheArgSerGlyTrp-----AspLysAsnAlaThrValM 413
Db 1577 TTAAGAGAGGG-----AGTAATGGTTATATAGGATATAAAGGCTACCGTA- 1623
QY 413 etValLeuLysAlaSerProGlyGlu-----PheHisAlaGlnP 427
Db 1624 -----AACTCACCACCTTCTCAAAATATCGTGTAGAGTTCATAGCCCACTA 1672
QY 427 roAspAsnGlyThrPheGluLeuPheIleLysGlyArgAsnPheThrProAspAlaGlyV 447
Db 1673 GTGTTTCTGGACTATTCAACGCTTTATT----- 1701
QY 447 alpHeValTyrSerGlyAspGluAlaIleMetLysLeuArgAsnTrpTyrArgGlnThra 467
Db 1701 ----- 1701
QY 467 rgIleHisSerThrLeuThrLeuAspAsnGlnAsnMetValIleThrLysAlaArgGlnA 487
Db 1702 -----AATGATGAATAGCGCTTCAAAAAAATTTCAAA 1735
QY 487 snLysTrpGluThr-----GlyAsnAsnLeuAspValLeuThrTyr----- 500
Db 1736 GTACTTAGAACAATAAGTGAAGAAAGAT---TTAACCTATGGTTTCATTTGGATATA 1792
QY 501 -----ThrAsnProSerTyrProAsn-----LeuA 509
Db 1793 TAGAATATTCTACGACCAITCAATTTCCGAATGAGCATCCAAAATCACTCTTCATTAA 1852
QY 509 spHis-----GlnArgSerValLeuPheIleAsnLysLysTyrPheLeuValIleAspA 527
Db 1853 ACCATTGTGATAACAATTCACCAATTTATAGTATCAATCGAATTTATCCCTGTAGAT- 1911
QY 527 rgAlaIleGlyGluAlaThrGlyAsnLeuGlyValHisTrpGlnLeuLysGluAspSerA 547
Db 1912 -----GTAAATTATGATGAAGAAAAA--- 1935
QY 547 snProValPheAspLysThrLysAsnArgValTyrThrThrTyrArgAspGlyAsnL 567
Db 1936 -----CTAGAAAAAGCACAGAAAGCCGTGAATACCTTGTTTACAGAGGAGAAATG 1987
QY 567 euMet-----IleGlnSerLeuAsnAlaAspArgThrSerLeu 579
Db 1988 CACTCCAAAATAAGTGACAGATTATAAGTGACAGGTTTCAATT 2034

RESULT 7
US-08-328-962-1
; Sequence 1, Application US/08328962
; Patent No. 5536637
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; TITLE OF INVENTION: A NOVEL METHOD FOR ISOLATING CYTOKINES
; TITLE OF INVENTION: AND OTHER SECRETED PROTEINS
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Legal Affairs, Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,962
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/08/045,267
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: McDaniels, Patricia A.
; REGISTRATION NUMBER: 33,194
; REFERENCE/DOCKET NUMBER: GI 5200
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-876-1170
; TELEFAX: 617-876-5851
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1542 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Saccharomyces cerevisiae
; IMMEDIATE SOURCE:
; CLONE: SUC2
; US-08-328-962-1

Alignment Scores:
Pred. No.: 0.000208 Length: 1542
Score: 123.00 Matches: 109
Percent Similarity: 32.01% Conservative: 76
Best Local Similarity: 18.86% Mismatches: 179
Query Match: 3.52% Indels: 214
DB: 1 Gaps: 31

US-08-802-285A-2 (1-659) x US-08-328-962-1 (1-1542)
QY 94 valThrArgGluMetAlaAspLysAlaLeuValHisGlnPheGlnProHisLysGlyTyr 113
Db 4 ATGACAAACGAAATAGCATAGACCTTTGGTCCAC---TTCACATCCCAACAAAGGGCTGG 60
QY 114 -----GlyTyrPheAspTyrGlyLysAspIleAsnTrpGlnMetTrp--- 127
Db 61 ATCAATGACCCAAATGGTGTGGTACGATGAAAGATGCCAAATGGCATCTGTACTTT 120
QY 128 ProValLysAspAsnGluValArgTrpGlnLeuHisArgValLysTrpTrpGlnAlaMet 147
Db 121 CAATACACCCAAATGACACCGTATGG---GGTACGCCATTGTTTGGGCG--- 168
QY 148 AlaLeuValTyrHisAlaThrGlyAspGluLysTyrAlaArgGluTrpValTyrGlnTyr 167
Db 169 -----CATGTACTCCGATGAT-----TTG 189
QY 168 SerAspTrpAlaArgLysAsnProLeuGlyLeuSerGlnAspAsnAspLysPheValTrp 187
Db 190 ACTAATGG---GAAGATCAACCCATTGCTATCGCTCCCAAGCGTAAC----- 234
QY 188 ArgProLeuGluValSerAspArgValGlnSerLeuProProThrPheSerLeuPheVal 207
Db 234 ----- 234
QY 208 AsnSerProAlaPheThrProAlaPheLeuMetGluPhe----- 220
Db 235 GATTCAGGTGCTTTCTCTGCTCCATGGTGGTGGTATTACAAACACAGATGGGTGTTTC 294
QY 221 -----LeuAsnSerTyrHis 225
Db 295 AATGATACTATTGATCCAGACAAAGATGGTGGATTTGGACTTATACACTCTCTGAA 354
QY 226 GlnGlnAlaAspTyrLeuSerThrHisTyrAlaGluGlnGlyAsnHisArgLeuPheGlu 245
Db 355 AGTGAAGAGCAATACATTAGC-----TATTCTCTTGTGGTGGTGTACACTTTTACTGAA 408
QY 246 AlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPheLysAsp----- 261
Db 409 TACCAAAAGAACCCCTGTTTAGCTGCCAATCCCACTCAATTCAGAGATCCAAAGGTGTTTC 468
QY 262 -----SerProArgTrpArgGlnThrGlyIleSerValLeuAsnThrGluIle 277
```

```
Db 469 TGGTATGAACCTTCTCAAAATGGATTATGACG-----GCTCCCAAAATCACAAGACTAC 522
QY 278 LysLysGlnValTyrAlaAsp----- 284
Db 523 AAAATTGAAATTTACTCTCTCTGAGTCTGAAGTCTGGAAGCTAGAATCTGCAATTGGC 582
QY 285 -----GlyMetGlnPheGluLeu----- 290
Db 583 AACGAGGTTTCTTAGCTACCAATACGATGTCAGGTTTATTGAAGTCCCACTGAG 642
QY 291 -----SerProIleTyrHisValAlaAlaAspIle----- 301
Db 643 CAAGATCCTTCCAAATCTTATTGGTCTGTTATTCTTATCAACCCAGGTGCACCTGCT 702
QY 302 -----PheLeuLysAlaTyr 306
Db 703 GCGGTTCTTCAACCAATATTGTTGGTATGCTTCAATGGTACTCAATTTGAAGCGTTT 762
QY 307 GlySerAlaLysArgValAsnLeuGluLysGluPheProGlnSerTyr-----ValGln 324
Db 763 GACAATCAATCTAGAGTG-----GTAGATTGTTGTAAGACTACTATGCTTGCA 813
QY 325 ThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuProAspTyrAsnThrPro 344
Db 814 ACTTTCTTCAAC-----ACTGACCAACCTACGTTTCAGCA 849
QY 345 MetPheGlyAspSerTrpIleThrAspLysAsnPheArgMetAlaGlnPheAla----- 362
Db 850 TTA-----GGTATTGCTTGGCTTCA-----AACTGGGAGTACAGTGCCTTTGCCAACT 900
QY 363 -----SerTrpAlaArgValPheProAlaAsnGlnAlaIleLys 375
Db 901 AACCCATGAGATCATCGATGCTTTGGTCCGCAAGTTTCTTTGAC-----ACTGAA 954
QY 376 TyrPheAla-----ThrAspGlyLysGlnGlyLysAlaProAsnPheLeuSerLys 392
Db 955 TATCAAGCTAATCCAGAGACTGAATTGATCAATTTGAAAGCCGCAACCAATATTGAAC--- 1011
QY 393 AlaLeuSerAsnAlaGlyPheTyrThrPheArgSerGlyTrpAspLysAsnAlaThrVal 412
Db 1012 ---ATTAGTAATGCTGGT-----CCCTGGTCTGCTTTTCTACTAAC 1050
QY 413 MetValLeuLysAlaSerProGlyGluPheHisAlaGlnProAspAsnGlyThrPhe 432
Db 1051 ACAACTCTAATAAGGCCCAATTTCTTACAAATGCGATTGAGCAACTCGACTGCTACCTTA 1110
QY 433 GluLeuPheIleLysGlyArgAsnPheThrProAspAlaGlyValPheValTyrSerGly 452
Db 1111 GAGTTTGAGTTG-----GTTTACGCTGTTAACACC 1140
QY 453 AspGluAlaIleMetLys-----LeuArgAsnTrpTyrArgGlnThrArg 467
Db 1141 ACACAAACCATATCCAAATCCGCTCTTCCGCACTTATCATTTGGTTCAAGGGTTAGAA 1200
QY 468 -----IleHisSerThrLeuThrLeuAsp 475
Db 1201 GATCCTGAAGAATATTTGAGAATGGGTTTGAAGTCAGTGTCTTCTCTCTTTTGGAC 1260
QY 476 AsnGlnAsnMetValIleThrLysAlaArgGlnAsnLysTrpGluThrGly----- 492
Db 1261 CGTGTAACTCTAAGGTCAAGTTTGTCAAGGAGAACCCATATTTCCACAAACAGAAATGCT 1320
QY 493 ---AsnAsnLeuAspValLeuThrTyrThrAsnProSerTyrProAsn----- 507
Db 1321 GTCACAAACCAACCACTTCAAGTCTCAGAACCACTAAGTTACTATATAAGTGTACGGCCTA 1380
QY 508 LeuAspHisGlnArgSerValLeuPheIleAsn-----LysLys 520
Db 1381 CTGGATCAAAACATCTTGGAAATGTACTTCAACGATGGAGATGTGGTTTCTTACAAATACC 1440
QY 521 TyrPheLeuValIleAspArgAlaIleGlyGluAlaThrGlyAsnLeuGlyVal 538
Db ----- 538
```

```
Db 1441 TACTTTCATGACCACCGGTAAACGCTCTAGGATCTGTGAACATGACCACCTGGTGTC 1494
RESULT 8
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2
Alignment Scores:
Pred. No.: 185 Length: 4403765
Score: 120.50 Matches: 85
Percent Similarity: 34.2% Conservative: 60
Best Local Similarity: 20.0% Mismatches: 150
Query Match: 3.45% Indels: 129
DB: 4 Gaps: 22
US-09-802-285A-2 (1-659) x US-09-103-840A-2 (1-4403765)
QY 131 AspAsnGluValArgTrpGlnLeuHisArgValLysTrpTrpGlnAlaMetAla----- 148
Db 3679590 GACAAGGCTATCGAGTGG---CTGCACACCGGTGCGGCGCCAGCAATGCCACCAAGCGGTGG 3679534
QY 149 LeuValTyrHisAlaThrGlyAspGluLysTyrAlaArgGluTrpValTyrGlnTyrSer 168
Db 3679533 ATGCTGTACTACCGCACCGGC---GCCACCCACGCGCCACACCACTGATTCAAG----- 3679483
QY 169 AspTrpAlaArgLysAsnProLeuGlyLeuSerGlnAspAsnAspLysPheValTrpArg 188
Db 3679482 GAATGGCGCGCAACAGTACCAGGTGAGTTCGATGATGCTGGATGTGTACCGGCAGAG 3679423
QY 189 ProLeuGluValSerAspArgValGlnSerLeuProProThrPheSerLeuPheValAsn 208
Db 3679422 ACATTCGAACCGCAAAACGACTCGGATCATTCACCCGACGCGCAACTC---ACCGAG 3679366
QY 209 SerProAlaPheThrProAlaPheLeuMetGluPheLeuAsnSerTyrHisGlnGlnAla 228
Db 3679365 CGGCCCGACCTATTCCCGCGGTGG----- 3679342
QY 229 AspTyrLeuSerThrHisTyrAlaGluGlnGlyAsnHisArgLeuPheGluAlaGlnArg 248
Db 3679341 GACAGTATGTCG-----GAGGCGCAAAACCGCTCTTTGCCCGCCAGATG 3679297
QY 249 AsnLeuPheAlaGlyValSer----- 255
Db 3679296 GAGGTGTTCCGCGGTTTCTCGAAATGCGGACTGGAAATGTTGGCCGCTCTGACACGCG 3679237
QY 256 PheProGluPheLysAspSerProArg-----TrpArgGlnThrGly 269
Db 3679236 ATCGAGGATCTCGGCGAGTCCGACACACAGTGTGGTGTCTACATCTGGGGCGACAAATGGC 3679177
QY 270 IleSerValLeuAsnThr-----GluIleLysLysGlnValTyrAlaAspGlyMet 286
Db 3679176 GCCAGCATGGAGGCGCACCAACACCGGTTTCTGTTCAATGATGATGACGCTTCTTAAACGCGCTG 3679117
```



```
QY 287 -----GlnPheGluLeuSerProIleTyr---HisValAlaAlaIle 299
Db 3679116 GATCTGGATGCGAGCGGCAATTTGGAGCTGATCAACAATATTCGCGCGCATCGCGCACTC 3679057
QY 300 -----AspIlePheLeuLysAlaTyrGlySerAlaLysArgValAsn 313
Db 3679056 GCGCAGAGTTTACCCACCGCATTTTGGCAGCGCTGGCGCATCGGACACACCCCG 3678997
QY 314 LeuGlu-----LysGluPheProGlnSerTyrValGlnThrValGluAsnMetIleMet 331
Db 3678996 TTGCAGTGGGCAAGCAGATGGCCACCCATCGGCGGCGCATCCATTGGTGGTC 3678937
QY 332 Ala----- 332
Db 3678936 GCTTGGCCGCGCGGATGCGGCGAGCGCGGTTCGTAGCAGTTCCACCACCTGCATC 3678877
QY 333 -----LeuIleSerIleSerLeuProAsp----- 340
Db 3678876 GACATCGCGCGCGGATGTTGGCGGCATCGTTTACCGAGCGGACCCATGTCGAGCGC 3678817
QY 341 ---TyrAsnThrProMetPheGlyAspSerTrpIleThrAspLysAsnPheArgMetAla 359
Db 3678816 TTCGACGAGAACCGATGGACGGAACCACTTCGTG----- 3678781
QY 360 GlnPheAlaSerTrpAlaArgValPheProAlaAsnGlnAla-----Ile 374
Db 3678780 -----CGGACCTTCGACGAGCTGGAAGCGCGAAGACCGCCACACCGTG 3678739
QY 375 LysTyrPheAlaThrAspGlyLysGlnGlyLysAlaProAsnPheLeuSerLysAlaLeu 394
Db 3678738 CAGTACTTC-----GAAACTTCGCGAGCGGTGCATC 3678706
QY 395 SerAsnAlaGlyPheTyrThrPhe-----ArgSerGlyTrpAspLysAsnAla 410
Db 3678705 TACAAAGACGCGTGGTGGCGTGGCTGCTGGACAGCGCGCTGGATCTGTCAACG 3678646
QY 411 ThrValMetValLeuLysAlaSerProProGlyGluPheHisAlaGlnProAspAsnGly 430
Db 3678645 GAGACAGATG-----CGACGGTTTCGCGCGCGGACCTAC-----GACCGGACCGAGAC 3678598
QY 431 ThrPheGluLeuPheIleLysGlyArgAsnPheThrProAspAlaGlyValPheValTyr 450
Db 3678597 GTCGGAGCTGTACTACCTACAGATGACTTCTCCAGCGGAAACCTGGCAGCGGAG 3678538
QY 451 SerGlyAspAlaIleMetLysLeuArgAsnTrpTyrArgGlnThrArgIleHisSer 470
Db 3678537 CATCCCGACAAGTCCGCGAGCTCACCAGCTGCTGGTGGCAGGAGCGGACGAAACCGG 3678478
QY 471 ThrLeuThrLeu 474
Db 3678477 GTGCTCGCGCTG 3678466

RESULT 9
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis

; OTHER INFORMATION: H37Rv
US-09-103-840A-1
Alignment Scores:
Pred. No.: 186 Length: 4411529
Score: 120.50 Matches: 85
Percent Similarity: 34.20% Conservative: 60
Best Local Similarity: 20.05% Mismatches: 150
Query Match: 3.45% Indels: 129
DB: 4 Gaps: 22

US-09-802-285A-2 (1-659) x US-09-103-840A-1 (1-4411529)
QY 131 AspAsnGluValArgTrpGlnLeuHisArgValLysTrpGlnAlaMetAla----- 148
Db 3684804 GACAAGGCTATCGAGTGG---CTGCACACCGTGGCGGCGCCAGTAATCCCAAGCGGTGG 3684748
QY 149 LeuValTyrHisAlaThrGlyAspGluLysTyrAlaArgGluTrpValTyrGlnTyrSer 168
Db 3684747 ATGCTGTACTACGCGACCGGC---GCCACCCAGCGCCACACACGATATTCAG----- 3684697
QY 169 AspTrpAlaArgLysAsnProLeuGlyLeuSerGlnAspAsnAspLysPheValTrpArg 188
Db 3684696 GAATGGCGCGACAAGTACCGAGTGAGTTCGATGATGGCTGGATGTGTACCGCGAGAAG 3684637
QY 189 ProLeuGluValSerAspArgValGlnSerLeuProProThrPheSerLeuPheValAsn 208
Db 3684636 ACATTGGAAGCGCAAAAGCGACTCGGGATCAITCCACCGCGACCGCAACTC---ACCGAG 3684580
QY 209 SerProAlaPheThrProAlaPheLeuMetGluPheLeuAsnSerTyrHisGlnAla 228
Db 3684579 CGGCCGAGCTATTCGCCGCTGG----- 3684556
QY 229 AspTyrLeuSerThrHisTyrAlaGluGlnGlyAsnHisArgLeuPheGluAlaGlnArg 248
Db 3684555 GACAGTATCTCG-----GAGCGCAAAACCGCTCTTTCGCCCGCAGATG 3684511
QY 249 AsnLeuPheAlaGlyValSer----- 255
Db 3684510 GAGTGTTCGCCGGGTCTCGGAAATCGGACATGGATGTTGGCGCGGTGTGGACGCG 3684451
QY 256 PheProGluPheLysAspSerProArg-----TrpArgGlnThrGly 269
Db 3684450 ATCGAGGATCTCGGCGAGTCCGACACACAGTGGTGTCTTACATCTGGCGCGCATGGC 3684391
QY 270 IleSerValLeuAsnThr-----GluIleLysLysGlnValTyrAlaAspGlyMet 286
Db 3684390 GCCAGCATGGAGGGCAACCAACACCGGTTCTGTTCAATGAGATGACGTTCTTAAACGCGCTG 3684331
QY 287 -----GlnPheGluLeuSerProIleTyr---HisValAlaAlaIle 299
Db 3684330 GATCTGGATGCGGAGCGCAATTTGGAGCTGATCGAACAATACGCGCGCATCGCGCACTC 3684271
QY 300 -----AspIlePheLeuLysAlaTyrGlySerAlaLysArgValAsn 313
Db 3684270 GCGCAGAGTTTACCCACCGCATTTTGGCAGCGCTGGCGCATCGGACACACCCCG 3684211
QY 314 LeuGlu-----LysGluPheProGlnSerTyrValGlnThrValGluAsnMetIleMet 331
Db 3684210 TTGCAGTGGGCAAGCAGATGGCCACCCATCGGCGGCGCATCCATTGGTGGTC 3684151
QY 332 Ala----- 332
Db 3684150 GCTTGGCCGCGCGGATGCGGCGAGCGCGGTTCGTAGCAGTTCCACCACCTGCATC 3684091
QY 333 -----LeuIleSerIleSerLeuProAsp----- 340
Db 3684090 GACATCGCGCGCGGATGTTGGCGGCATCGTTTACCGAGCGGACCCATGTCGAGCGC 3684031
QY 341 ---TyrAsnThrProMetPheGlyAspSerTrpIleThrAspLysAsnPheArgMetAla 359
Db 3684030 TTCGACGAGAACCGATGGACGGAACCACTTCGTG----- 3683995
```


QY 292 ProIleTyrHisValAlaIleAspIlePheLeuLysAlaTyrGlySerAlaLysArg 311
Db 460 CATTTCCAAACAGATGACAAATGCTATTT-----TGTCAGAGAGAACACACGGA 510
QY 312 ValAsnLeuGluLysGluPheProGlnSerTyrValGlnThrValGluAsnMetIle--- 330
Db 511 CITATGCTAGAT---CAACACACACGCTTGTTACTCTACTACAGACCAATTAATTTCT 567
QY 331 -----MetAlaLeuIleSerIleSerLeuProAspTyrAsn 342
Db 568 GTAAGAGAATCTTTAAAGATGCTCGAATGCAAAATGCAACAGCTTTACACATTTTAAA 627
QY 343 ThrProMetPheGlyAspSerIleThrAspLysAsnPheArgMetAlaGlnPheAla 362
Db 628 AATCCATGCTATCC----- 642
QY 363 SerTTPAlaArgValPheProAlaAsnGlnAlaIleLysTyrPheAlaThrAspGlyLys 382
Db 643 -----ACAAATGATTAACACTCCAGATTTTAAATAGCGACACTGTC 681
QY 383 GlnGlyLysAlaProAsnPheLeuSerLysAlaLeuSerAsn-----AlaGly 398
Db 682 TTATTAACTTACCAATGATATACACAGCGGACTTTAAATCTTATATTACATCAAGG 741
QY 399 PheTyrThrPheArgSerGlyTTPAspLys-----AsnAla 410
Db 742 TATATTCAATTCGCAAGAGATGGAATCTGTAAATTATGATGAAAGTTTATAATCAA 801
QY 411 ThrValMetValLeuLysLysSerProGlyGluPheHisAlaGlnProAspAsnGly 430
Db 802 ACAAAGTTGATTTGCAAGCTGCTATT-----CAGGACTATTCTACTGTATCTACC 855
QY 431 ThrPheGluLeuPheIleLysGlyArgAsnPheThrProAspAlaGlyValPheValTyr 450
Db 856 ACTTTTGAATAATTCAAACTCTCTAAAT-----CCA 888
QY 451 SerGlyAspGluAlaIleMetLysLeuArgAsnTyrArgGlnThrArgIleHisSer 470
Db 889 TCAATAAAGATCTGTTAATAGTATATAGATATGTTGTTCCATGACTCTTCAATCT 948
QY 471 Thr-----LeuThrLeuAspAsn----- 476
Db 949 TTAGACATTCGTCACATGCGCTACTTTAGATATGTTAATTAATCCCTTCCAAATGATAG 1008
QY 477 -----GlnAsnMetValIleThrLysAlaArgGlnAsnLysTTPGlu----- 490
Db 1009 ATTCAATTCGATCAAACTCGCTTAGTATTTTCAGATGTTGCAGGACCTTGGGAAGTAAAT 1068
QY 491 -----ThrGlyAsnAsnLeuAspValLeuThrTyrThrAsnPro----- 503
Db 1069 GATAATATAACTTCGAATATTAGATGATTATACACCAATAATACAGGATAGATTT 1128
QY 504 -----SerTyrProAsnLeuAspHisGln----- 511
Db 1129 CAAGAAAGTTCAGATCTTAGAAATTTCACTTATCCAGTAAGATTAACAAGCATGCAA 1188
QY 512 -----ArgSerVal-----LeuPhe 516
Db 1189 TTCCATGGCAATATGTAACCTCAAAAGTGTAGAACATTTATAGCGATGCTCTTAA 1248
QY 517 IleAsnLysLysTyrPheLeuValIleAspArgAlaIleGlyGluAlaThrGlyAsnLeu 536
Db 1249 TTAATATTATAA-----AATAAACTATAACTGCAGGTGTAAGTAAATATT 1293
QY 537 GlyValHisTTPGlnLeuLysGluAspSerAsn----- 547
Db 1294 -----GATGAAGTAAATCAAAATAATAAACAATAACTATGCTCT 1332
QY 548 -----ProValPheAspLysThrLysAsnArgValTyrThrTyrArgAsp 563
Db 1333 GTAATAAATAGTCTTACTGATATCAAGTAAATTCACAAATTTCTCAATATTAGAT 1392

QY 564 GlyAsnAsnLeuMetIleGln-----SerLeu 572
Db 1393 TTAATATTAGTCAATGTAATAATGGTGTCAAAAGTAGCGGGTGTTCACACTTAGTTCA 1452
QY 573 AsnAlaAspArgThrSerLeuAsnGluGluGlyLysValSerTyrValTyrAsnLys 592
Db 1453 AATGGTAATTTCAATAATGCTGTTTACCTAATCAAAATAAATGTTATTATTTCAGTA 1512
QY 593 GluLeuLysArgProAlaPheValPheGluLysProLysLysAsnAlaGlyThrGln--- 611
Db 1513 CAATCAAAAT-----GATAAACCCAGAAAAACATCGACACACTTATAGA 1554
QY 612 -----AsnPheValSerIleValTyrProTyrAsp----- 621
Db 1555 AATGGGATATATAGACGAGTCAATTCCTTAATGATCTTGTCCAGAAAATGTAATGGA 1614
QY 622 -----GlyGlnLysAlaProGluIleSerIleArgGluAsnLysGlyAsnAsp 637
Db 1615 GATATAGATCCGGATACTAAACACCGTCATGCTCTCTT-----AAAGGGTTTCCG 1665
QY 638 PheGluLysGly 641
Db 1666 GCAGAAAAAGGA 1677
RESULT 12
PCT-US96-05320A-637
; Sequence 637, Application PC/TUS9605320A
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences
; APPLICANT: 9410 Key West Avenue
; APPLICANT: Rockville, MD 20850
; APPLICANT: United States of America
; APPLICANT: Johns Hopkins University
; APPLICANT: 720 Rutland Avenue
; APPLICANT: Baltimore, MD 21205
; APPLICANT: United States of America
; APPLICANT: Mark D. Adams
; APPLICANT: Owen White
; APPLICANT: Hamilton O. Smith
; APPLICANT: J. Craig Venter
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome,
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20003-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/05320A
; FILING DATE: April 22, 1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,102
; FILING DATE: June 7, 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,429
; FILING DATE: June 7, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Eric K. Steffe
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.014PC01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 637:
; SEQUENCE CHARACTERISTICS:

Db 2140 ATTGATAAAGTTCT-----GCAATAGATATCCGTTTATCAAAACCAAAAT 2187
QY 624 -----LysAlaProGluIleSerIleArgGluAsnLysGlyAsnAsp 637
Db 2188 CGAGATCAACGACGCTCAGGATATCGAAATGCTTCACGCTTGGAATGGGTGATTG 2247
QY 638 PheGluLys 640
Db 2248 TTTGAAAA 2256

RESULT 13
US-08-961-527-74
; Sequence 74, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340PI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 74:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16535 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-74

Alignment Scores:
Pred. No.: 0.0874 Length: 16535
Score: 115.00 Matches: 161
Percent Similarity: 33.03% Conservative: 94
Best Local Similarity: 20.85% Mismatches: 263
Query Match: 3.29% Indels: 254
DB: 4 Gaps: 45

US-09-802-285a-2 (1-659) x US-08-961-527-74 (1-16535)

QY 11 ValPheAlaValIleAlaLeuSerSerGly-----AsnIleLeuAla 24
Db 1387 GTAGTTGGAACCTGCTCTTGAAGAAAGGGAAGGAACTTGAAGAAACAACTCTAGAC 1446
QY 25 GlnSerSerIleThrArgLysAspPheAspHisIleAsnLeuGluTyrSerGlyLeu 44
Db 1447 AGCAAAACAAACTCGAATCGAATTCAGATTTCCTGGCTACTATTATCAATACCAA---ATC 1503
QY 45 GlnLysValAsnLysAlaValAlaAlaGlyAsnTyrAspAlaAlaLysAlaLeuLeu 64
Db 1504 GAGCGTCAGGTAAACACTGTTCTTGA-----CTCGATCCTTACGCTAAATCTCTTGCT 1557

QY 65 AlaTyrTyrArgGluLysSerLysAlaArgGluProAspPheSerAsnAlaGluLysPro 84
Db 1558 GCTTGAATAGCAGCATTTCCAAG-----1581
QY 85 AlaAspIleArgGlnProIleAspLysValThrArgGluMetAlaAspLysAlaLeuVal 104
Db 1582 -----ATTGACGAT---GCCATAAAGTGGCTAAAGCCGCTTTGTA 1620
QY 105 HisGlnPheGlnProHisLys---GlyTyrGlyTyrPheAspTyrGlyLysAspIleAsn 123
Db 1621 -----GATCCAGCTAAACTCGACCTCAAGACTTCAAGCTTATGCTAAGATTCACAAT 1671
QY 124 TrpGlnMetTrpProValLysAspAsnGluValArgTrpGlnLeuHisArgValLysTrp 143
Db 1672 TTCAG-----ACTCGTGAAGACGCCGCTTACTCTCAAGACTCAT-----1710
QY 144 TrpGlnAlaMetAlaLeuValTyrHisAlaThrGlyAspGluLysTyrAlaArgGluTrp 163
Db 1711 -----GTGCGTGATTTCACCTTCCAGATCCTGCCATTGCAAAAGACTTG 1752
QY 164 ValTyrGlnTyrSerAspTrpAlaArgLysAsnProLeuGlyLysSerGlnAspAsnAsp 183
Db 1753 ACC-----AAACCATTTGGGACTTTT-----GAA 1776
QY 184 LysPheValTrpArgProLeuGluValSerAsp-----ArgValGlnSerLeu 199
Db 1777 GCCTTCATTTGAAAAAAGTAGACTATCTCAAGACTTGGGTGTAAACCATATCAGCTCCTT 1836
QY 200 ProProThrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGlu 219
Db 1837 CCAGTCTTGCTTACTACTTCTTCAATGAA-----1866
QY 220 PheLeuAsnSerTyrHisGlnGlnAlaAspTyrLeuSerThrHisTyrAlaGluGlnGly 239
Db 1867 -----TTGAAAAACCATGAACGCTTGCTGACTACGCTTCAAGCAAC-----AGC 1911
QY 240 AsnHisArgLeuPheGluAlaGlnArgAsnLeuPheAla-----GlyValSerPhePro 257
Db 1912 AACTACAACTGGGGATATGACCCCTCAAAACTACTTCTCCTGCTGCTGATGATCTCAAGC 1971
QY 258 GluPheLysAspSerProArgTrpArgGlnThrGlyIleSerValLeuAsnThrGluIle 277
Db 1972 GATCCTAAG---AATCCAGAAAAACGAATCGCAGAAATTTAAAAACCTCATCAAGAAATC 2028
QY 278 LysLysGlnValTyrAlaAspGlyMetGlnPheGluLeuSerProIleTyr---HisVal 296
Db 2029 CACAAACGCT-----GGTATGGAGCTATCTTAGATGTCGTTTATAACCAACACA 2076
QY 297 AlaAlaIleAspIle-----PheLeuLysAlaTyr 306
Db 2077 GCCAAAGTCGATCTCTTTGAAGATTGGNACCAAACTACTACCATTTATGATGCCGAT 2136
QY 307 GlySerAlaLysArgValAsnLeuLysGluPheProGlnSerTyrValGlnThrVal 326
Db 2137 GGCACACCTCGA-----ACTAGCTTTGTGTGGAGCGCTTGGGGACAAC 2181
QY 327 GluAsnMetIleMetAlaLeuIleSerIleSerLeuPro-----339
Db 2182 CACCATATGACCAAAACGGCTCTTAATGACTCTATCAAACTACTAGTAGTACCTACAAA 2241
QY 339 -----339
Db 2242 GTGGATGGCTTCGTTTCGATATGATGGAGACCATGACCGCTTCTTATCGAAGAGCT 2301
QY 340 -----AspTyrAsnThrProMetPheGlyAspSerTrpIleThr 352
Db 2302 TACAAAGGTGCACGCGCCCTCAATCAAAACCTCATCATGCTTGGTGAAGTTGGAGAACC 2361
QY 353 -----AspLysAsnPhe---ArgMetAlaGlnPheAlaSerTrpAlaArgValPhe 368
Db 2362 TATGCGGTGATGAATAACATGCCTACTAAAGCTGTGACCAAGATTTGGATGAAA-----2415

```
QY 369 ProAlaAsnGlnAlaIleLysTyrPheAlaThrAspGlyLysGlnGlyLysAlaProAsn 388
Db 2416 ---CATACGATCTGCTGCTGCTTTTCAGATGACATCCGTT---AAC 2457
QY 389 PheLeuSerLysAlaLeuSerAsnAlaGlyPheTyrThrPheArgSerGlyTyrAspLys 408
Db 2458 AACCTCAAAATCTGGTTATCCAAACGAAGGTCAACCTGCCTTTATCACAGGTGGCAAGCGT 2517
QY 409 AsnAlaThrValMetVal-----LeuLysAlaSerPro 419
Db 2518 GATGTCACACCATCTTTAAATCTCATGCTCAACCACTAACTTTGAAGTGCAGC 2577
QY 420 ProGly-----GluPheHisAlaGlnProAspAsnGlyThr---PheGluLeuPhe 435
Db 2578 CTTGGAGATGTCATCCATATACCGCAGCCCATGATACTTGACCTCTTTGACATCAT 2637
QY 436 Ile-----LysGlyArgAsnPhe----- 441
Db 2638 GCCCAGTCTCAAAAAGACCAAGCAAGGCTGAGAACTATGCTGAAATCCACCGTCTG 2697
QY 442 -----ThrProAspAlaGlyValPheValTyrSerGly 452
Db 2698 TTACGACTGGAAATCTCATGCTTTCACAGCTCAAGAACTCCATTTATCCACTCCGCT 2757
QY 453 AspGluAlaIleMetLysLeuArgAsnTyrTyr-----ArgGlnThrArgIleHis 469
Db 2758 CAGGAA-----TATGGAGTACTAAACAATCCGTGACCCA 2793
QY 470 SerThrLeuThrLeuAspAsnGlnAsnMetValIleThrLysAlaArgGlnAsnLysTrp 489
Db 2794 GCCTCAAGACTCCAGTAGCAGAGATAAGGTTCCAAACAAATCTCATTTGTTGGTGAT 2853
QY 490 GluThrGlyAsnAsnLeuAspValLeuThrTyrThrAsnProSerTyrProAsnLeuAsp 509
Db 2854 AAGGACGGCAACCAATTTGACTATCTCTTACTTATCATGACTCTTACGATTTAGTGAT 2913
QY 510 HisGlnArgSerValLeuPheIleAsnLysLysTyrPheLeuValIleAspArgAlaIle 529
Db 2914 -----GCAGTCAACAAG----- 2925
QY 530 GlyGluAlaThrGlyAsnLeuGlyValHisTrpGlnLeuLysGluAspSerAsnProVal 549
Db 2926 -----TTTGACTGGACTAAGGCTACAGATGCTAAAGCTTAT 2961
QY 550 PheAspLysThrLysAsnArgValTyrThrTyrArgAspGlyAsnAsnLeuMetIle 569
Db 2962 CCTGAAATGTCAGAGCCGCT-----GACTATATGAAAGGTTTGAATGCGCTTCGT 3012
QY 570 GlnSerLeuAsnAla-----AspArgThrSerLeu 579
Db 3013 CAATCTACAGATGCTTCCGACTTAGAGTCTTCAAGATATCAAGACCGGTGCCACCTC 3072
QY 580 AsnGluGluGluGly-----LysValSerTyrValTyrAsnLysGluLeu 594
Db 3073 ATCACTGTCCAGGCCAAATGTTGGAAAAGAGGATGTAGTATTGGCTACCAATC 3132
QY 595 LysArgPro-----AlaPheVal-----PheGluLysProLysLys 606
Db 3133 ACTGCTCCAAACGGCGATATCTACGCGACTCTTTGTCAATCGGATGAAAAGCTCGGAA 3192
QY 607 ---AsnAlaGlyThr-----GlnAsnPheValSerIleValTyrProTyr 620
Db 3193 TTAAATTTGGAACTGCTTTCACATCTAAGAATTCGGAAGTTTGGCA----- 3243
QY 621 AspGlyGlnLysAlaProGluIleSerIleArgGluAsnLysGlyAsnAspPhe---Glu 639
Db 3244 GATGAAAACCAAGCAGCAGCAGTGGGAATTCGCAACCCGAAAGGACTTTGAATGACTGAA 3303
QY 640 LysGly---LysLeuAsn---LeuThrLeuThrIle 649
Db 3304 AAAGCTTGAATTTGAATGATGCCCTTACAGCTACTGTT 3339
```

RESULT 14

```
US-08-485-568A-5
; Sequence 5, Application US/08485568A
; Patent No. 5589382
; GENERAL INFORMATION:
; APPLICANT: Payne, Jewel
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Fu, Jenny M.
; TITLE OF INVENTION: No. 5589382el Bacillus thuringiensis Genes Encoding
; TITLE OF INVENTION: Nematode-Active Toxins
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,568A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/310,197
; FILING DATE: 21-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/918,345
; FILING DATE: 21-JUL-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/558,738
; FILING DATE: 27-JUL-1990
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/357,698
; FILING DATE: 16-DEC-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/176,403
; FILING DATE: 30-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/999,053
; FILING DATE: 14-JUL-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: MA48DD2.C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3504 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-485-568A-5
Alignment Scores:
Pred. No.: 0.00789
Score: 114.50
Percent Similarity: 31.49%
Best Local Similarity: 19.66%
Query Match: 3.28%
DB: 1
Length: 3504
Matches: 103
Conservative: 62
Mismatches: 170
Indels: 189
Gaps: 24
```

[illegible]


```
Qy 622 -----GlyGlnLysAlaProGluIleSerIleArgGluAsnLysGlyAsnAsp 637
Db 1615 GATATAGATCCGGATACCTAAACAACCGTCATTGCTTCTT-----AAAGGGTTTCCG 1665
Qy 638 PheGluLysGly 641
Db 1666 GCAGAAAAAAGGA 1677
```

Search completed: August 4, 2003, 14:14:52
Job time : 6299 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 4, 2003, 07:41:04 ; Search time 405 Seconds
(without alignments)
3664.363 Million cell updates/sec

Title: US-09-802-285A-2
Perfect score: 3494
Sequence: 1 MTTFKFKLIVFAVIALSSG.....KQKLNLTITNGKQQLVLP 659

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q/cg2 1/USPTO spool/US9802285/runat 28072003 190510 9692/app query.fasta 1.839
-DB=N Geneseq 101002 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -NXPEN=2000000000
-USER=US9802285 @CGN 1.1 263 @runat 28072003 190510 9692 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N Geneseq 101002: *
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT: *
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT: *
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT: *
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT: *
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT: *
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT: *
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT: *
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT: *
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT: *
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT: *
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT: *
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT: *
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT: *
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT: *
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT: *
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT: *
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT: *
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT: *
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT: *
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT: *
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT: *
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT: *
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT: *
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
1	3494	100.0	1980	17	AAQ99227	Heparinase-III gen
2	250.5	7.2	1902	24	ABN67223	Streptococcus poly
3	250.5	7.2	2155561	24	ABN71527	Streptococcus poly
c	234.5	6.7	1923	24	ABN67224	Streptococcus pneu
5	176	5.0	3763	19	AAV52319	Enterococcus faeca
c	166.5	4.8	11427	20	AAI13102	Listeria innocua D
7	139	4.0	2295	24	ABQ69654	Listeria innocua D
c	139	4.0	3011208	24	ABQ69245	Streptococcus poly
9	137.5	3.9	3495	24	ABN70606	Streptococcus poly
10	137.5	3.9	3522	24	ABN67343	Streptococcus epi
11	132.5	3.8	2097	24	ABN91332	Staphylococcus cere
12	127	3.6	3994	5	AAH40009	Saccharomyces cere
13	127	3.6	3994	12	AAQ11883	Invertase gene. S
14	123.5	3.5	3414	14	AAQ41744	DNA encoding B.t.
15	123	3.5	1542	19	AAV09317	Unsecreted yeast i
16	122.5	3.5	3753	21	AAA05807	Group B Streptococ
17	121	3.5	3414	15	AAQ56921	Bacillus thuringie
18	120.5	3.4	3756	24	ABN67342	Streptococcus poly
19	120.5	3.4	3759	24	ABN70380	Streptococcus poly
c	120.5	3.4	4403765	22	AAI99683	Mycobacterium tube
c	120.5	3.4	4411529	22	AAI99682	Mycobacterium tube
22	118.5	3.4	3414	15	AAQ56922	Bacillus thuringie
23	118	3.4	2319	17	AAQ99226	Heparinase-II gene
24	118	3.4	7787	24	ABQ71070	Listeria monocytog
25	117.5	3.4	4455	18	AAV25095	H. pylori secreted
26	117.5	3.4	8709	22	AAF25628	H. pylori HPN165 e
27	117.5	3.4	2944528	24	ABA03041	Listeria monocytog
28	116	3.3	2215	21	AAZ43884	C. vicina LSP-2 CD
29	115.5	3.3	3414	14	AAQ41745	DNA encoding B.t.
30	115.5	3.3	3504	18	AAI62498	Nematocidal toxin
31	115.5	3.3	1163020	24	ABQ67197	Listeria innocua C
32	115.5	3.3	3011208	24	ABQ69245	Listeria innocua D
33	115	3.3	3222	24	ABK75363	Bacillus lichenifo
34	115	3.3	16535	19	AAV52207	Streptococcus pneu
c	115	3.3	92584	21	AAF22288	BAC containing rep
c	115	3.3	1830121	17	AAI42063	Haemophilus influe
37	114.5	3.3	2859	19	AAV42656	Human adenovirus 5
38	114.5	3.3	3504	18	AAI89185	Nematode toxin 167
39	114.5	3.3	3504	18	AAI61018	Bacillus thuringie
40	114.5	3.3	3903	19	AAI14480	H. pylori GHP0 109
41	114.5	3.3	4785	14	AAQ48797	B. obbensis Trp (188
42	114.5	3.3	1830121	17	AAI42063	Haemophilus influe
43	113.5	3.2	3845	20	AAV59397	Avian topoisomeras
44	113.5	3.2	4765	17	AAI08132	Protease gene. Py
45	113.5	3.2	4765	18	AAI85670	Pyrococcus furiosu

ALIGNMENTS

RESULT 1

AAQ99227

ID AAQ99227 standard; DNA; 1980 BP.

XX AAQ99227;

XX AAQ99227;

DT 07-APR-1996 (first entry)

XX Heparinase-III gene.

DE Heparinase-III gene.

XX Heparinase-III; heparan sulphate degradation; PCR; primer; cloning;

KW Escherichia coli; polymerase chain reaction; ribosome binding site;

KW toxicity; vector; haemostatic; blood-clotting; antibody;

KW affinity chromatography; ds.

XX

OS Flavobacterium heparinum.

XX

XX Key

XX Location/Qualifiers

EH

sig_peptide 1..72
 mat_peptide /*tag= a
 /*tag= b

FT 1..72
 FT /*tag= a
 FT 73..1977
 FT /*tag= b
 XX
 PN WO9534635-A1.
 XX
 PD 21-DEC-1995.
 XX
 PF 09-JUN-1995; 95WO-US07391.
 XX
 XX 10-JUN-1994; 94US-0258639.
 PR
 XX (IBEX-) IBEX TECHNOLOGIES.
 PA (ZIMM/) ZIMMERMANN J.
 XX
 PI Bennett C, Blain F, Gu K, Musil R, Su H, Zimmermann J;
 XX WPI; 1996-097381/10.
 DR P-PSDB; AAR89265.
 XX

Nucleic acids encoding Flavobacterium heparinum heparinase III and
 III - for degrading heparin and heparan sulphate, also related host
 cells, proteins and antibodies, useful in heparinase purification.

Claim 8; Fig 8; 7spp; English.

The sequence encodes Flavobacterium heparinum heparinase-III
 (EC-4.2.2.8), which degrades heparan sulphate. The sequence is
 isolated by polymerase chain reaction amplification using
 e.g. primers 3-1 to 3-4 (AAQ99234-099237) or degenerate primers 3-5
 to 3-8 (AAQ99238-099241). Toxicity of the gene and natural selection
 of the host against clones with the entire sequence has been
 circumvented by cloning sections of the gene separately in
 Escherichia coli using a vector with a modified ribosome binding
 site, which increases expression levels. The heparinase-III may be
 used to neutralise anticoagulant activity. Antibodies against the
 protein may be used to differentiate between native and recombinant
 enzymes, and when immobilised they may be used for heparinase
 purification by affinity chromatography.

SQ Sequence 1980 BP; 629 A; 432 C; 439 G; 480 T; 0 other;

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
0	3494.00	1980	659
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	17	Gaps:	0

US-09-802-285A-2 (1-659) x AAQ99227 (1-1980)

QY	1	MetThrThrLysIlePheLysArgIleIleValPheAlaValIleAlaLeuSerSerGly	20
DB	1	ATGACTACGAAATTTTAAAGAGGATCATTTGCTGCTGTAATTCGCCCTATCGTCGGGA	60
QY	21	AsnIleLeuAlaGlnSerSerIleThrArgLysAspPheAspHisIleAsnLeuGlu	40
DB	61	AATATACTTGCACAAAGCTCTTCCATTACACGAGAAAGATTGTGACCATCAACCTTGAG	120
QY	41	TyrSerGlyLeuGluLysValAsnLysAlaValAlaGlyAsnTyrAspAlaAla	60
DB	121	TATTCGGACTGGAAAGGTTAATAAAGCAGTTGCTCGCGCAACTATGACGATCGCGCC	180
QY	61	LysAlaLeuLeuAlaTyrTyrArgGluLysSerLysAlaArgGluProAspPheSerAsn	80
DB	181	AAAGCATTTACTGGCATACTACAGGGAAGAAAGTAAAGCCAGGAACTGATTTTCAGTAA	240
QY	81	AlaGluLysProAlaAspIleArgGlnProIleAspLysValThrArgGluMetAlaAsp	100
DB	241	GCAGAAAGCGCTGCGGATATACGCCAGCCCATAGATAGGTTACGCGTGAATGGCCAC	300

QY	101	LysAlaLeuValHisGlnPheGlnProHisLysGlyTyrGlyTyrPheAspTyrGlyLys	120
DB	301	AAGGCTTTGGTCCACAGTTTCAACCGCAAGGCTACGGCTATTTTTCATTATGTGTA	360
QY	121	AspIleAsnTrpGlnMetTrpProValLysAspAsnGluValArgTrpGlnLeuHisArg	140
DB	361	GACATCAACTGGCAGATGTGGCGGTAAAGACAACTGAAGTACGCTGGCAGTTGCACCT	420
QY	141	ValLysTrpTrpGlnAlaMetAlaLeuValTyrHisAlaThrGlyAspGluLysTyrAla	160
DB	421	GTAAATGGTGGCAGGCTATGGCCCTGGTTTATCAGCTACGGCGGATGAAATATGCA	480
QY	161	ArgGluTrpValTyrGlnTyrSerAspTrpAlaArgLysAsnProLeuGlyLeuSerGln	180
DB	481	AGAGATGGGTATATCAGTACAGCGATTGGGCCAGAAAGAACCCATTGGGCTGTGCGAG	540
QY	181	AspAsnAspLysPheValTrpArgProLeuGluValSerAspArgValGlnSerLeuPro	200
DB	541	GATAATGATAAATTTGTGTGGCGGCCCTTGAAGTGTGGACAGGCTACAAAGTCTTCCC	600
QY	201	ProThrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGluPhe	220
DB	601	CCAACTTCAGCTTATTTGTAACCTGCGCAGCCTTTACCCAGCCTTTTAAATGGAATTT	660
QY	221	LeuAsnSerTyrHisGlnGlnAlaAspTyrLeuSerThrHisTyrAlaGluGlnGlyAsn	240
DB	661	TTAAACAGTTACCAACCAACAGGCCGCTTATTTATCTAGCATTTATGCCAAGGGAAC	720
QY	241	HisArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPheLys	260
DB	721	CACCGTTTATTTGAAGCCCAACGCAACTGTTTTCAGGGGTATCTTCCCTGAATTTAA	780
QY	261	AspSerProArgTrpArgGlnThrGlyLysSerValLeuAsnThrGluLysLysGln	280
DB	781	GATTCCACCAAGATGGAGCCAAACCGGCATATCGGTCTGAACCCAGATCAAAAGACAG	840
QY	281	ValTyrAlaAspGlyMetGlnPheGluLeuSerProIleTyrHisValAlaAlaIleAsp	300
DB	841	GTTTATGCGGATGGATGCGAGTTTGAACCTTTCACCAATTTACCATGTAGTCCCATCG	900
QY	301	IlePheLeuLysAlaTyrGlySerAlaLysArgValAsnLeuGluLysGluPheProGln	320
DB	901	ATCTTCTTAAAGGCCATATGTTTGGAGATTTCAGGATTACAGATAAAATTTTCAGATG	360
QY	321	SerTyrValGlnThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuProAsp	340
DB	961	TCTTATGTACAAACTGTAGAAATATGATTATGGCGCTGATCAGTATTTTCATGCGCAG	1020
QY	341	TyrAsnThrProMetPheGlyAspSerTrpIleThrAspLysAsnPheArgMetAlaGln	360
DB	1021	TATAACACCCCTATGTTTGGAGATTTCAGGATTACAGATAAAATTTTCAGATGCGCAG	1080
QY	361	PheAlaSerTrpAlaArgValPheProAlaAsnGlnAlaIleLysTyrPheAlaThrAsp	380
DB	1081	TTTGCGAGCTGGCGCGCGGTTTCCCGCAACACCGCCATATAAATATTTTGTACAGAT	1140
QY	381	GlyLysGlnGlyLysAlaProAsnPheLeuSerLysAlaLeuSerAsnAlaGlyPheTyr	400
DB	1141	GGCAAAACAGTAAGGCGCCCTAACTTTTATCCAAAGCATTTGAGCAATTCAGAGCTTT	1200
QY	401	ThrPheArgSerGlyTrpAspLysAsnAlaThrValMetValLeuLysAlaSerProPro	420
DB	1201	ACGTTTGAAGCGGATGGGATAAAATGCAACCGTTATGTTATGTTTAAAGCCAGTCTTCC	1260
QY	421	GlyGluPheHisAlaGlnProAspAsnGlyThrPheGluLeuPheIleLysGlyArgAsn	440
DB	1261	GGGAAATTTTCATGCCCGCGGATTAACGGGACTTTTGAACCTTTTATAAAGCGGAGAAC	1320
QY	441	PheThrProAspAlaGlyValPheValTyrSerGlyAspGluAlaIleMetLysLeuArg	460
DB	1321	TTTACCCAGACGCGCGGGGTTATTTGTATATGCGGCGAGGAGCATCATGAAATCGCG	1380
QY	461	AsnTrpTyrArgGlnThrArgIleHisSerThrLeuThrLeuAspAsnGlnAsnMetVal	480

```

Db 1381 AACTGGTACGTCNAACCCGCATACACAGACGCTTACATCGACATCAAAATAGGTC 1440
Qy 481 ILeThrIysAlaArgGlnAsnLysTyrGluThrGlyAsnAsnLeuAspValLeuThrTyr 500
Db 1441 ATTACCAAGCCGGCAAAACAAATGGGAAACAGGAATTAACCTTGATGTGCTTACCTAT 1500
Qy 501 ThrAsnProSerTyrProAsnLeuAspHisGlnArgSerValLeuPheLeuAsnLysLys 520
Db 1501 ACCAACCCAGCTATCCGAATCTGGACCATCAGCGCAGTGTACTTTTCATCAACAAAAA 1560
Qy 521 TyrPheLeuValIleAspArgAlaIleGlyGluAlaThrGlyAsnLeuGlyValHisTyr 540
Db 1561 TACTTCTGTCTCGATAGGGCAATAGGCGAAGCTACCGAAACCTGGCGGTACACTGG 1620
Qy 541 GlnLeuLysGluAspSerAsnProValPheAspLysThrLysAsnArgValTyrThr 560
Db 1621 CAGCTTAAAGACAGACCAACCCCTGTTCGATAAGACAAAGAACCGGGTTTACCCACT 1680
Qy 561 TyrArgAspGlyAsnAsnLeuMetIleGlnSerLeuAsnAlaAspArgThrSerLeuAsn 580
Db 1681 TACAGAGATGGTACACACCTGATGATCCAAATCGTTGAATCGGCACAGGACAGCCTCAAT 1740
Qy 581 GluGluGluGlyLysValSerTyrValTyrAsnLysGluLeuLysArgProAlaPheVal 600
Db 1741 GAAGAGAAGAAAGGTATCTTATGTTTACAATAAGGAGCTGAAAGACCTGCTTCGTA 1800
Qy 601 PheGluLysProLysLysAsnAlaGlyThrGlnAsnPheValSerIleValTyrProTyr 620
Db 1801 TTTGAAAGCCTAAAAAGAAATGCGGCACACAAATTTTGTCAATAGTATAGTTATCCATAC 1860
Qy 621 AspGlyGlnLysAlaProGluIleSerIleArgGluAsnLysGlyAsnAspPheGluLys 640
Db 1861 GAGCGCCAGAGGCTCAGAGATCAGATACGCGGAAACCAAGGCAATGTTTGGAAA 1920
Qy 641 GlyLysLeuAsnLeuThrLeuThrIleAsnGlyLysGlnGlnLeuValLeuValPro 659
Db 1921 GGCAAGCTTAATCTAACCTTACCATTAAACGAAACAAACAGCTTGTTGGTTCCT 1977

RESULT 2
ID ABN67223
ID ABN67223 standard; DNA; 1902 BP.
AC AC
AC ABN67223;
XX
DT 01-JUL-2002 (first entry)
DE Streptococcus polynucleotide SEQ ID NO 2359.
XX
KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
OS Streptococcus agalactiae.
XX
XX W0200234771-A2.
XX
PD 02-MAY-2002.
XX
XX 29-OCT-2001; 2001WO-GB04789.
XX
XX 27-OCT-2000; 2000GB-0026333.
XX
XX 24-NOV-2000; 2000GB-0028727.
XX
XX 07-MAR-2001; 2001GB-0005640.
XX
XX (CHIR-) CHIRON SPA.
XX
XX (GENO-) INST GENOMIC RES.
XX
XX Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
XX Tettelin H;
XX
XX WPI; 2002-352536/38.
XX
XX P-PSDB; ABP26592.

```

```

XX
PT New Streptococcus protein for the treatment or prevention of infection
PT or disease caused by Streptococcus bacteria, such as meningitis, and
XX for detecting a compound that binds to the protein -
XX
PS Claim 7; Page 3387-3388; 4525pp; English.
XX
CC The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins.
XX
SQ Sequence 1902 BP; 641 A; 264 C; 362 G; 635 T; 0 other;

Alignment Scores:
Pred. No.: 1,82e-14 Length: 1902
Score: 250.50 Matches: 127
Percent Similarity: 36.09% Conservative: 74
Best Local Similarity: 22.80% Mismatches: 223
Query Match: 7.17% Indels: 133
DB: 24 Gaps: 21

US-09-802-285A-2 (1-659) x ABN67223 (1-1902)
Qy 123 AsnTyrGlnMet-----Tyr-----Tyr----- 127
Db 145 AATTGGGATATGGAACCTTGTCTAAAGCATATGTCTTGCATCTTTGGATGGGATAAG 204
Qy 128 ProValLysAspAsn---GluValArgTyrGlnLeuHisArgValLysTyrTyrGlnAla 146
Db 205 CCAATACAGATGATCCAGATGGTGTATATGCTAAATCGTCAAACTTATCTTTTAAA 264
Qy 147 MetAlaLeuValTyrHisAlaThrGlyAspGluLysTyrAlaArgGluTyrValTyrGln 166
Db 265 TTCCTGGTGTCTACATTGTTGAAGCGATAAGCTTACCTTAGACAAATGAAATATTTT 324
Qy 167 TyrSerAspTrpAla-----ArgLysAsnProLeuGlyLeuSerGlnAspAsn 182
Db 325 ATGTACCATTTGGATTTGATTTCAATTTACACTAAAGCCAGAGGGCGG----- 372
Qy 183 AspLysPheValTyrArgProLeuGluValSerAspArgValGlnSerLeuProProThr 202
Db 373 -----GTAAGTCGTACTATTGACACTGGGATACGTTGCATGTCATGGCTCAAAGTT 423
Qy 203 PheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGluPheLeuAsn 222
Db 424 TTGATTTTTTTAGATTATTTTGATTAAATAACAGAACTAAAAAATAAATACTATTAACT 483
Qy 223 SerTyrHisGlnGlnAlaAspTyrLeuSerThrHisTyrAlaGluGln-----Gly 239
Db 484 AGTTTACGGGAGCAGATAAATCTTATATGAGGAGGATATTTATCGTGAAGAAGATAGTCTAAGT 543
Qy 240 AsnHisArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPhe 259
Db 544 AACTGGGAATTTTACAACACACAGCAATATTTGGCGTGTATTACTATTATGAGAGATGAA 603
Qy 260 LysAspSerProArgTyrArgGlnThrGlyIleSerValLeuAsnThrGluIleLysLys 279
Db 604 TTAATCTACCTCAATTCAGAGTGTTCGTGGAAGAGAAATTATTACTTCAATCAAGCTT 663

```



```

QY 123 AsnTrpGlnMet-----Trr----- 127
D 1845827 AATTGGGATAGAACCTTGTCTAAAGCATATTGCTTGATCCTTTGGAGTGGGATAAG 1845768
QY 128 ProValIysAspAsn-----GluValArgTrpGlnLeuHisArgValIysTrpTrpGlnAla 146
D 1845767 CCAGTAACAGATGATCAGAAATGGTTGATATGCTAAATCGTCAAACTTATCTTTTAAA 1845708
QY 147 MetAlaLeuValTyrHisAlaThrGlyAspGluIysTyrAlaArgGluTrpValTyrGln 166
D 1845707 TTCTGTGTGTCTACATTGTTGAAGGCGATAAAGCTTACCTTAGACAAATGAATATTTT 1845648
QY 167 TyrSerAspTrpAla-----ArgIysAsnProLeuGlyLeuSerGlnAspAsn 182
D 1845647 ATGTACCATTGGATTGATGTCATTTACACTAAAGCCAGAGGGCG----- 1845600
QY 183 AspIysPheValTrpArgProLeuGluValSerAspArgValGlnSerLeuProProThr 202
D 1845599 -----GTAAGTCGTACTATTGACACTGGGATACGTTGCATGTCATGGCTCAAAAGTT 1845549
QY 203 PheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGluPheLeuAsn 222
D 1845548 TTGATTTTTTTAGATATTATTTGGATTAAACAGAAACTAAAAAATTTAACTATTAACT 1845489
QY 223 SerTyrHisGlnGlnAlaAspTyrLeuSerThrHisTyrAlaGluGln-----Gly 239
D 1845488 AGTTACGGGACGACATACTTATATGAGGACTATTATTCTGTGAAAAAGATAGTCTAAGT 1845429
QY 240 AsnHisArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPhe 259
D 1845428 AACTGGGGAATTTTACAAACAAACAGCAATATTGCGGTGTTTATACTATTATGAAGATGAA 1845369
QY 260 LysAspSerProArgTrpArgGlnThrGlyLeuSerValLeuAsnThrGluIleLysLys 279
D 1845368 TTTAACTACTCGAAATTCAGAGTTTGTGTAAGAGAAATTTATTTACTTCAAACTCAAGCTT 1845309
QY 280 GlnValTyrAlaAspGlyMetGlnPheGluLeuSerProIleTyrHisValAlaAlaIle 299
D 1845308 CAAATTTAGATGACGGGACCCAGTATGAACATCAATTAATGATCATGATAGAAGTCTTG 1845249
QY 300 AspIlePheLeuIysAlaTyrGlySerAlaLysArgValAsnLeuGluLysGluPhePro 319
D 1845248 AAGTCCTTAATGGAACATGATATCTTGGCGCTAAATATTATCTACCATTAGAA----- 1845195
QY 320 GlnSerTyrValGlnThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuPro 339
D 1845194 -----GAACTATTGAAAAATGTTATCTTGGCGCTAAATATTATCTACCATTAGAA 1845147
QY 340 AspTyrAsnThrProMetPheGlyAspSerTrpIleThrAspIysAsnPheArgMetAla 359
D 1845146 GATTATTGTCACTAGCTATAGGGACAGTCAATGACCGACTACTCGT----- 1845099
QY 360 GlnPheAlaSerTrpAlaArgValPheProAlaAsnGlnAlaIleLysTyrPheAlaThr 379
D 1845098 GATATTCTAACTTTGGCAACACTTGTTGTTGAAATCTCTTAAAAACAAATCATTTCTTTT 1845039
QY 380 AspGlyLysGln-----GlyLysAlaProAsnPheLeuSerLys 392
D 1845038 GATAAGTTAATTAGAACTTTACTTTGTTGGGAAGCCATCAATTTATCTTTTGA 1844979
QY 393 AlaLeuSerAsnAla-----GlyPheTyrThrPheArgSerGlyTyrAsp--- 407
D 1844978 GAAATACCGCGTGGACATAGGAGTCTGCTTATCTTTTCCAGATTCGTGTCATGTG 1844919
QY 408 -----LysAsnAlaThrValMetValLeuLysAlaSerProGlyGluPheHis 424
D 1844918 TGTTTACGTGATGATAGCGGTATATATTTTAAAGAACGTCATTTGTTAGCGCTCAT 1844859
QY 425 AlaGlnProAsnGlyThrPheGluLeuPheIleLysGlyArgAsnPheThrProAsp 444
D 1844858 ACTCATAGTATATAATATGTTGTTCTCTATGATAAAAGAACCTATTTTCAITGAT 1844799
QY 445 AlaGlyValPheValTyrSerGlyAspGluAlaIleMetLysLeuArgAsnTyrTyrArg 464

```

```

D 1844798 GCAGGAGATATATCTTCAAAAGAACAA-----CTAAGTATGATTTTAAA 1844751
QY 465 GlnThrArgIleHisSerThrLeuThrLeuAspAsnGlnAsnMetValIleThrLysAla 484
D 1844750 CGTTCGACTAGTCATTCACATGATGACCTTGATGGGCAACCCTTAGAAATGATCAAG-- 1844694
QY 485 ArgGlnAsnLysTrpGluThrGlyAsnAsnLeuAspValLeuThrTyrThrAsnProSer 504
D 1844693 -----GACTCTTTGGACATACAAT-----TCT 1844673
QY 505 TyrProAsnLeuAspHis----- 510
D 1844672 TATCCAAAATGTGACTATTGTCAGTTGCTCAAGATAGGTACCATTTAGTCGAAGGA 1844613
QY 511 -----GlnArgSerValLeuPheIleAsnLysLysTyrPheLeuValIle 525
D 1844612 CAACTACATGTCCAAAGAGCTTCTGATATCTATTACCATAAGCGATGTTGTTAACTTTA 1844553
QY 526 AspArgAlaIleGlyGluAlaThrGlyAsnLeuGlyValHisTrpGlnLeuLysGluAsp 545
D 1844552 CCGAGGCCATT----- 1844541
QY 546 SerAsnProValPheAspLysThrLys-----AsnArgValTyrThr- 560
D 1844540 ---ACCTTAGTTATGATAGGTGAGTTGCCAGGAGAGCATGCTTAAACAATCAATAT 1844484
QY 561 -----TyrArgAspGly-----AsnAsnLeuMetIleGlnSer 571
D 1844483 ATTTTAGATGATCAGGTCATTTATGAAATGGGTTTGTAAATGACTTGAATTAGTAAGT 1844424
QY 572 LeuAsnAlaAspArgThrSerLeuAsnGluGluGlyLysValSerTyrValTyrAsn 591
D 1844423 -----CCTACGACCTTTAATCTAGAAGATTGCTCTTAAAGCGGTATAAT 1844376
QY 592 -----LysGluLeuLysArgProAlaPheValPheGlu 602
D 1844375 CAATTGCAGAAAAGTCATAAATAGTTAAGAAAATAAAATTTGTTGATGAG 1844325
RESULT 4
ID ABN67224
XX ABN67224 standard; DNA; 1923 BP.
XX AC ABN67224;
XX DT 01-JUL-2002 (first entry)
XX DE Streptococcus polynucleotide SEQ ID NO 2361.
XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
XX KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
XX OS antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX PN Streptococcus pyogenes.
XX PP WO200234771-A2.
XX PD 02-MAY-2002.
XX PF 29-OCT-2001; 2001WO-GB04789.
XX PR 27-OCT-2000; 2000GB-0026333.
XX PR 24-NOV-2000; 2000GB-0028727.
XX PR 07-MAR-2001; 2001GB-0005640.
XX PA (CHIR-) CHIRON SPA.
XX PA (GENO-) INST GENOMIC RES.
XX PI Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
XX PI Tettelin H;
XX DR WPI; 2002-352536/38.
XX DR P-PSDB; ABP26593.

```


Db 1572 C---TGCCTTTCACATCATCAAGAAACGATTTTGTCAAAAAATACACCAAACTCTGAC 1628
 QY 595 sArgProAlaPheValPheGluLysProLysLysAsnAlaGlyThrGlnAsnPheValSe 615
 Db 1629 TAGTCACAAACTCATCAAAACGAAACCCCTTTAAAGACAGGGTTGCACCACT-----AC 1682
 QY 615 rIleValTyProTyAspGlyGlnLysAlaProGluIleSerIleArgGluAsnLysG1 635
 Db 1683 CCTATTAGTCCCTGACATACCAAGTTACTCTTAAACACCACTTCAACACAGGCAACG 1742
 QY 635 yAsnAspPheGluLysGlyLysLeuAsnLeuThrIleAsnGlyLysGln 653
 Db 1743 CAACCAATAGACTGCC-----CTTTCTGCATTTAAAGGAAGCAG 1788

RESULT 5

AAV52319/c

ID AAV52319 standard; DNA; 3763 BP.

XX AAV52319;

XX 23-OCT-1998 (first entry)

DT Streptococcus pneumoniae genome fragment SEQ ID NO:186.

DE Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;

KW computer readable medium; vaccine; pharmaceutical composition; ds.

XX Streptococcus pneumoniae.

XX WO9818931-A2.

XX 07-MAY-1998.

XX 30-OCT-1997; 97WO-US19588.

XX 31-OCT-1996; 96US-0029960.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;

PI Kunsch CA, Rosen CA;

XX WPI; 1998-272225/24.

XX Computer-readable medium with recorded Streptococcus pneumoniae

PT polynucleotide sequences - useful in diagnostic kits and assays, and

PT pharmaceutical compositions and vaccines for Streptococcus

PT pneumoniae

XX Claim 1; Page 1135-1137; 1409pp; English.

PS The present invention describes a computer readable medium which has
 CC the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
 CC recorded on it, or a representative fragment or a sequence at least 95%
 CC identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in
 CC SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
 CC Streptococcus pneumoniae. The present invention also describes an
 CC isolated nucleic acid molecule encoding a homologue of any of the
 CC fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the
 CC nucleic acid molecule is produced by a process comprising: (a) screening
 CC a genomic DNA library using as a probe a target sequence defined by any
 CC of the sequences in SEQ ID NO:1 to 391, identifying members of the
 CC library which contain sequences that hybridize to the target sequence and
 CC isolating the nucleic acid molecules from the members; or (b) isolating
 CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
 CC molecules whose nucleotide sequence is homologous to amplification
 CC primers derived from the fragment of the S. pneumoniae genome to prime
 CC the amplification and isolating the amplified sequences. The computer
 CC readable medium can be used in a computer-based system for identifying
 CC fragments of the S. pneumoniae genome of commercial importance, or
 CC expression modulating fragments of the S. pneumoniae genome. Products
 CC from the present invention can be used in diagnosis kits and assays, and

CC pharmaceutical compositions and vaccines for S. pneumoniae.

XX Sequence 3763 BP; 1178 A; 804 C; 677 G; 1104 T; 0 other;
 SQ

Alignment Scores:

Pred. No.: 8.69e-07 Length: 3763
 Score: 176.00 Matches: 116
 Percent Similarity: 34.63% Conservative: 80
 Best Local Similarity: 20.49% Mismatches: 231
 Query Match: 5.04% Indels: 140
 DB: 19 Gaps: 20

US-09-802-285A-2 (1-659) x AAV52319 (1-3763)

QY 71 SerLysAlaArgGluProAspPheSerAsnAlaGluLysProAlaAspIleArgGlnPro 90
 Db 2173 TCAAAAGACTATGAGAAAGTCAAGAGTCCCTTGAACGCTTGATGCAACATCGTTTATG 2114
 QY 91 IleAspLysValThrArgGluMetAlaAspLysAlaLeuValHisGlnPheGlnProHis 110
 Db 2113 TTTGATAGT---CCTTGGGATATGGAGCCTTGTTCAAAATCCATCAATCCAGCCG--- 2060
 QY 111 LysGlyTyrGlyTyrPheAspTyrGlyLysAspIleAsnTrpGlnMetTrpProValLys 130
 Db 2059 -----ATGGTATGGGATCAAGTAITTTGAAGAT 2033
 QY 131 AspAsnGluValArgTrpGlnLeuHisArgValLysTrpTrpGlnAlaMetAlaLeuVal 150
 Db 2032 GATCCAGAATGGTCTTATATGCTCAATCAGACAGATATCTCTTCGATGTTATGATAGGG 1973
 QY 151 TyrHisAlaThrGlyAspGluLysTyrAlaArg-----GluTrp 163
 Db 1972 TATCTGGTAGAGAGATAGGACTAATTCAAAAGTGCAGTCTCTTCTATTGATGG 1913
 QY 164 ValTyrGlnTyrSerAspTrpAlaArgLysAsnProLeuGlyLeuSerGlnAspAsnAsp 183
 Db 1912 ATTGACGAGGTGAGAGAAATTTCTCCTCAATCCTTGATGACTAGAACCTTGGATACGGGT 1853
 QY 184 -----LysPheValTrpArgProLeuGluValSerAspArgValGlnSerLeuProPro 201
 Db 1852 ATTCGTTCTCTTACTTGGTTGAAACTA-----1826
 QY 202 ThrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGluPheLeu 221
 Db 1825 ---CTCTTCTCTCTTGTAAATTTGACTTGTAGGAGAGAAAGAACTAGAGAAATTTTG 1769
 QY 222 AsnSerTyrHisGlnGlnAlaAspTyrLeuSerThrHisTyrAlaGluGln-----238
 Db 1768 GTCAGTCTAGAAAAGCAGATTGACTTTATGAAAAGCTACTATCGCGCCCAAGTACACCCCT 1709
 QY 239 GlyAsnHisArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGlu 258
 Db 1708 AGTAACCTGGGGGATTTTACAAAACAATTCGCGTCTCTATCTATCTTCTTTCAGAT 1649
 QY 259 PheLysAspSerProArgTrpArgGlnThrGlyIleSerValLeuAsnThrGluLeuLys 278
 Db 1648 AGATGGATCTAGAGAACCTTACCATTTGCTTCAGAGAGTTGAAAACAGCAATTTGAG 1589
 QY 279 LysGlnValTyrAlaAspGlyMetGlnPheGluLeuSerProIleTyrHisVal-----296
 Db 1588 ACACAGATTTTAGGAGATGGAAGCCAGTTTGAACAGTTCGATTCTCTATCATGTAGAGTT 1529
 QY 297 -----AlaAlaIleAspIlePheLeuLysAlaTyrGlySerAlaLysArgValAsnLeu 314
 Db 1528 TATAAAGCCTCTGCGATTGTGTCTC-----TTG 1499
 QY 315 GluLysGluPheProGlnSerTyrValGlnThrValGluAsnMetIleMetAlaLeuIle 334
 Db 1498 CTTCCAGATTCGACAGATAGTTACCAAGAGTTGCTGGAAAAGATGGCGCACCTATATCAA 1439
 QY 335 SerIleSerLeuProAspTyrAsnThrProMetPheGlyAspSerTrpIleThrAspLys 354
 Db 1438 ATGATCAGAGGCTTAGTACGACGGACTTTGGCTTTTGGTGATAGCGATTCTACAGAACG 1379

```
QY 355 AsnPheArgMetAlaGlnPheAla----- 362
Db 1378 ACAGAAATTTGAGCCTGCTCTGTGGTTTGAACGAGAACCTTCTTAAACGGTCTG 1319
QY 363 -----SerTrpAlaArgValPheProAlaAsnGlnAlaIleLysTyrPhe 377
Db 1318 GATGTTAAAGTTGATTTGCTTAGCCTCTTGTCTGGCGGAGAAAAGGTCAAGCGACTG 1259
QY 378 AlaThrAspGlyGlnGlyLysAlaProAsnPheLeuSerLysAlaLeuSerAla 397
Db 1258 CAGGAATTTGAAAAGAGAGCTTGGCAGCCTAAG-----TCCATGATCTTTGAAGACTCT 1205
QY 398 GlyPheTyrThrPheArgSerGlyTrpAspLysAenAlaThrValMetValLeuLysAla 417
Db 1204 GGACATGCTGCATTAAGAT-----GAACATCGTATCTATTATTTTCAAAAT 1157
QY 418 SerProGlyGluPheHisAlaGlnProAspAsnGlyThrPheGluLeuPheIleLys 437
Db 1156 GGTCCGCTAGGAAGTGCCTATAGTCATAGCAGAGATAGTTTTTTCCTTACAGTATCAA 1097
QY 438 GlyArgAsnPheThrProAspAlaGlyValPheValTyrSerGlyAspGluAlaIleMet 457
Db 1096 GGCCAACTTATTTTCATAGATGCTGGCGCTTATCTTATCG-GGAGATATATCAACGTTA 1038
QY 458 LysLeuArgAsnTrpTyrArgGlnThrArgIleHisSerThrLeuThrLeuAspAsnGln 477
Db 1037 TCTCTTAAAGAGTCTTGGAGTCATTCGACCTGCATGT----- 999
QY 478 AsnMetValIleThrLysAlaArgGlnAsnLysTrpGlu---ThrGlyAsnAsnLeuAsp 496
Db 998 -----AGATGGAAAGTCCGGAAGAAGATCACGGG 969
QY 497 ValLeu-ThrTyr----- 500
Db 968 ATCTCTGGGAATATGAATATCTCTCACTCCCTGTTTGTCCACCATAAAGAAGGAGGG 909
QY 501 -----ThrAsnProSerTyrProAsnLeuAspHi 510
Db 908 AGTGCATATATAGGGGGCTTATTTGGTCAGCAGAACCTGATTTGCCTTATCTT---CA 852
QY 510 sGlnArgSerValLeuPheIleAsnLysLysTyrPheLeuValIleAsp-----ArgAl 528
Db 851 CAGGAGAAAATCTCTCATGTTGTAGAGAGATGCTGGCTTGTGTAGATCATCATCAGGNG 792
QY 528 atleGlyGluAlaThrGlyAsnLeuGlyValHisTrpGlnLeuLysGluAspSerAsnPr 548
Db 791 TCAAGGTCAA-----CATGAGGTGTTGACTCAGTTT----- 761
QY 548 oValPheAspLysThrLysAsnArgValTyrThrTyrArgAspGly-----AsnAs 566
Db 760 -ATCCTTGACAGGAT-----GTGACCTATCAAGATGGGAAAATCAATCA 717
QY 566 nLeuMetIleGlnSer 571
Db 716 GTTGAGACTATGGAGT 701
```

RESULT 6

```
AA13102/c
ID AAX13102 standard; DNA; 11427 BP.
XX
AC AAX13102;
XX
DT 19-MAR-1999 (first entry)
XX
DE Enterococcus faecalis genome contig SEQ ID NO:165.
XX
KW Enterococcus faecalis; contig; detection; Enterococcal infection;
KW vaccine; attenuation; computer readable medium; ds.
OS Enterococcus faecalis.
XX
PN WC9850555-A2.
```

```
XX 12-NOV-1998.
XX
PF 04-MAY-1998; 98WO-US08985.
XX
PR 14-NOV-1997; 97US-0066009.
PR 06-MAY-1997; 97US-0044031.
PR 16-MAY-1997; 97US-0046655.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Barash SC, Dillon PJ, Kunsch CA;
XX
XX WPI; 1999-045171/04.
DR
XX
PT New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
XX
PS Claim 1; Page 932-938; 2084pp; English.
XX
CC A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC AAX12938 to AAX13919 represent these nucleotide sequences which are
CC primary nucleotide sequences, also known as contigs. The computer-based
CC system can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
XX
SQ Sequence 11427 BP; 3451 A; 2545 C; 1839 G; 3587 T; 5 other;
```

Alignment Scores:

Pred. No.:	3,378-05	Length:	11427
Score:	166.50	Matches:	121
Percent Similarity:	35.18%	Conservative:	89
Best Local Similarity:	20.27%	Mismatches:	215
Query Match:	4.77%	Indels:	174
DB:	20	Gaps:	26

US-09-802-285A-2 (1-659) x AAX13102 (1-11427)

```
QY 120 LysAspIleAsnTrpGlnMetTrpProValLysAspAsnGluValArgTrpGlnLeuHis 139
Db 7871 AAAGATATGCTGGAATCGGTATCTGATGATGATCCAGATGCTCTTTATGTTAGT 7812
QY 140 ArgValLysTrpTrpGlnAlaMetAlaLeuValTyrHisAlaThrGlyAspGlyLysTyr 159
Db 7811 CGTCAAGCTTTCTTGTAGATCTGGCACAAGCATATGCATTTACTTAAAAAAGACGTTAC 7752
QY 160 AlaArgGluTrpValTyrGlnTyrSerAspTrpAlaArgLysAsnProLeuGlyLeuSer 179
Db 7751 TTACAGAAATGGCAGAGCTTGGCTGCTTAGATGTTGGATTCGATGATCAACAAAC--- 7698
QY 180 GlnAspAsnAspLysPheValTrpArgProLeuGluValSerAspArgValGlnSerLeu 199
Db 7697 AATTCGACGAATAGGATGTTTGGCTCGTGTAGATGTTGGATTCGATGATCAACAAAC--- 7641
QY 200 ProProThrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGlu 219
Db 7640 -----TGGATGAAAAGCTTGAACGTATATTTCCAAATCGCTGATTTTCA 7599
QY 220 PheLeu-----AsnSerTyrHisGlnGlnAlaAspTyrLeuSer 232
Db 7598 CTATTAGGAATTGATGATGTGTTGAACACAGCCCTGCTGATCATCTGACTATTGAG 7539
```


QY 523 LeuValIleAspArgAlaIleGlyGluAlaThrGlyAsnLeuGlyValHisTrpGlnLeu 542
 Db 1903 -----GCAAAAGATAATTTAAACTTACGAGGAGCA 1935
 QY 543 LysGluAspGlnSerProValPheAspLysThrLysAsnArgValTyrThrTyrArg 562
 Db 1936 AAATGGGATAGT-----GAAATTTCTCAAGTAGTAGAA 1968
 QY 563 AspGlyAsnAsnLeuMetIleGlnSerLeuAsnAlaAspArgThrSerLeuAsn----- 580
 Db 1969 AAAGGTGCAAAAGTAACATAATTAATTCG-----AAAACAAGTATAAATGGTTGG 2016
 QY 581 -----GluGluGluGlyValSerTyrValTyrAsnLysGluLeuLysArgPro 597
 Db 2017 CATCAAGTAACCTACGGTGGTAAATAAGGTTATTAIG-----ATTTTAAGTGAT 2064
 QY 598 AlaPheValPheGluLysPro 604
 Db 2065 AACTATTATTAGTCGAAACCA 2085

RESULT 8

ABQ69245/c

ID ABQ69245 standard; DNA; 3011208 BP.

XX AC ABQ69245;

XX XX 29-AUG-2002 (first entry)

XX DE Listeria innocua DNA sequence #694.

XX XX Antibacterial; Listeria; food contamination; mutational analysis;

XX KW infection; ds.

XX XX Listeria innocua.

XX OS WO200228891-A2.

XX PN 11-APR-2002.

XX XX 04-OCT-2001; 2001WO-FR03061.

XX XX 04-OCT-2000; 2000FR-0012697.

XX XX (INSP) INST PASTEUR.

XX PA (CNRS) CNRS CENT NAT RECH SCI.

XX XX Kunst F, Glaser P;

XX PI WPI; 2002-332479/37.

XX DR New genomic sequences from Listeria species, useful for detection,

XX PT treatment and prevention of infection, also related polypeptides,

XX PT antibodies and modulators -

XX PS Claim 5; SEQ ID 2058; 180pp; French.

XX XX The present invention relates to nucleic acid sequences

CC (ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes

CC and primers for identification and/or detection of Listeria (e.g. as

CC contaminants in foods, or mutational analysis) and for analysis of

CC gene expression. Proteins encoded by the nucleic acid sequences can be

CC used to screen for compounds that modulate gene expression, replication

CC and pathogenicity of Listeria (potential therapeutic agents), also for

CC treating infections by Listeria, and are useful as immunogens in

CC anti-Listeria vaccines.

CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic format

CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX XX

SQ Sequence 3011208 BP; 941651 A; 568176 C; 559189 G; 942192 T; 0 other;

Alignment Scores:

Pred. No.: 33.9 Length: 3011208
 Score: 139.00 Matches: 148
 Percent Similarity: 31.94% Conservative: 97
 Best Local Similarity: 19.30% Mismatches: 280
 Query Match: 3.98% Indels: 242
 DB: 24 Gaps: 37
 US-09-802-285A-2 (1-659) x ABQ69245 (1-3011208)
 QY 7 LysArgIleIle-----ValPheAlaValIle-----AlaLeuSerSerGly 20
 Db 2117803 AAAAGAAATTATATCTGTATTTGTCATTTTCTACTGGGATTTTAGCCCATTC 2117744
 QY 21 AsnIleLeuAlaGlnSerSerSerIleThrArgLysAspPheAspHisIleAsnLeuGlu 40
 Db 2117743 ACAATTAAGCAAGTGCAGCAACCAATACAAATAAATTTGAAGAGCTT----- 2117693
 QY 41 TyrSerGlyLeuGluLysValAsnLysAlaValAlaAlaGlyAsnTyrAspAspAlaAla 60
 Db 2117692 -----TATGCAAGAGCTAAA 2117678
 QY 61 LysAlaLeu---LeuAlaTyrTyrArgGluLysSerLysAlaArgGluProAspPheSer 79
 Db 2117677 AAACATTTAGGTAAACCATATTCGCAAGACAGGACAGAGAGAGAGGCGCCAAATTTACTTC 2117618
 QY 80 AsnAla-----GluLysProAlaAspIleArgGlnPro--- 90
 Db 2117617 GACTGCTCAGGTTACACACAATATGTTTATGAATAAGTAAACAGGAGTAAGAATTCCTAAT 2117558
 QY 91 -----IleAspLysValThrArgGluMetAlaAspLysAla 102
 Db 2117557 ACTTCAGCTCCTCAATATTTCAGCTGCTGATTAAGTGAAATGGTAATCAAAAACCTGGT 2117498
 QY 103 LeuValHisGlnPheGlnProHisLysGlyTyrGlyTyrPheAspTyrGlyLysAspIle 122
 Db 2117497 GATTTAGTATATTTCAAAGCTCATGTAGGAATA---TACATAGTAATGCTAAATGATT 2117441
 QY 123 AsnTrpGlnMetTrpProValLysAspAsnGluValArgTrpGlnLeuHisArgValLys 142
 Db 2117440 AATGCTCAAAATGATGGAGTCAAAATAGACAACATT-----AATAGTAGT 2117396
 QY 143 TrpTrpGlnAlaMetAlaLeuValTyr-----HisAlaThrGlyAspGluLysTyrAla 160
 Db 2117395 TATTCGAGAGTATTTTGTGTGATCGGAAGATTTTCAATTTTCCCGAAGAAAAAGGA 2117336
 QY 161 ArgGluTrpValTyrGlnTyrSerAsp-----Trp----- 170
 Db 2117335 TCTAAATCCGCTATGCTGTGTCAGATTTAAATTTACGATCTAGCAATAACTGGGATAGT 2117276
 QY 171 -----AlaArgLysAsnProLeuGly-----LeuSerGlnAspAsnAspLys 184
 Db 2117275 TCAGTAGCTGGAAGATTCACCAAGGCGCTAAAGTATCTATTGACATTAGACAGTGAAG 2117216
 QY 185 PheValTrpArgProLeuGluValSerAspArgValGlnSerLeuProProThrPheSer 204
 Db 2117215 AATGGTTGGTGTATGTTACTTATAATAATACTAAAGTTATATGTTAAATACACTAAT 2117156
 QY 205 LeuPheValAsnSerProAlaPheThrProAlaPhe----- 216
 Db 2117155 TATTTCTCGATACACCTGTGATAAACAACATCATATGCCAAGGATTAATAAACTTAAGA 2117096
 QY 216 ----- 216
 Db 2117095 ACGAAAGCAACATGGGATAGCGATGTAGCTCAAAAAGTACAAAAGGTGAAAAGTTACT 2117036
 QY 217 -----LeuMetGluPheLeuAsnSerTyrHisGlnGlnAlaAspTyrLeuSerThrHis 234
 Db 2117035 GTGAACCTTAAACCTAATGTGTAATGTTGTATCAAGTAAACATACACGCGGAAAAACAGGC 2116976
 QY 235 TyrAlaGluGlnGlyAsnHisArgLeuPheGluAlaGlnArgAsnLeu-----PheAla 252
 Db 2116975 TATATGATTTTAAATAATAATTACTTTAGTAGAATAATCCCTTGAATATGGAACGATTAT 2116916

QY 253 GlyValSerPheProGluPheLysAspSerProArgTyrArgGlnThrGlyIleSerVal 272
Db 2116915 GCGGTGGTACCTTAAATCGCTAGTCAGCAACTGG---CATAGTAGCATTTAGCTTAA 2116959
QY 273 LeuAsnThrGlu-----lleLysLysGlnValTyrAlaAspGlyMetGlnPheGlu 289
Db 2116858 GTGTACCAAGAGCGAGCTGTAAAGTAGAA-----ATGGATACAAAT 2116914
QY 290 LeuSerProIleTyrHisValala----- 297
Db 2116813 AGTGGTCCCTGGTATTAAGTAAATACATATCAAAACCAACAGGTGTACATCCCACTAACAGAT 2116754
QY 298 -----AlaIleAspIlePheLeuLysAlaTyrGlySerAlaLysArgValaAsnLeu 314
Db 2116753 GATTATTTATCAAAACTACTGTTCTAAAACGATATACGCTTAAGATATTTAAACCTTA 2116694
QY 315 GluLysGlu-----PheProGlnSerTyrValGlnThrValGlu-----AsnMetIle 330
Db 2116693 AGAACAAAGCAACTTTGGGATAGCGAGCTTGCTCAAAAAGTACAAAAGGTGAAAAGTA 2116634
QY 331 MetAlaLeuIleSerIleSerLeuProAspTyrAsnThrProMetPheGly----- 347
Db 2116633 ACAGTTAATTAAACCAAGTGTAAATGGTGTATCAAGTAACTTATGGTGTGTAATAAAA 2116574
QY 348 -----AspSerTrpIleThrAspLysAsnPheArgMetAlaGlnPhe 361
Db 2116573 GGTATCATGATCTTAATGATACACTATTAGTCGAAAAGCGCTTAATATGAACCTAT 2116514
QY 362 -----AlaSerTrp-----AlaArg 366
Db 2116513 TATCGGTGAGTAGCTTAAATTTACGTAGTGGCGAAATGGATAGTAGTAACTCA 2116454
QY 367 ValPheProAlaGlnAlaIleLys-----TyrPhe 377
Db 2116453 GTGGTTCAGAGGTAGCTGTCGAAAGTCAAAAGTCGAAATGACACGCAATGTTGGCAATCGTT 2116394
QY 378 AlaThr-----AspGlyLysGlnGlyLysAlaPro-----AsnPheLeuSerLys 392
Db 2116393 AAAGTAACCTATGACATAAACAAGGTATATGCCCTAAATGATTTATTTATCTGAA 2116334
QY 393 AlaLeuSerAsnAlaGlyPheTyrThr-----PheArgSerGly 405
Db 2116333 ACCGCTGTGTAAAAACCTATTATGCAAAAGATAATTTAACTAGTACGCAAGCAAAA 2116274
QY 406 TrpAspLysAsnAlaThrValMetValLeuLysAlaSerProProGlyGluPheHisAla 425
Db 2116273 TGGGATGACGAGTTACTCAAAAGTAGAAAAGGTGAAAAGTAAACAGTCAATTCGAA 2116214
QY 426 GlnProAspAsnGlyThrPheGluLeuPheIleLysGlyArgAsnPheThrProAspAla 445
Db 2116213 ACAAGTATCGATGGCTGGTATGAATG----- 2116187
QY 446 GlyValPheValTyrSerGlyAspGluAlaIleMetLysLeuArgAsnTrpTyrArgGln 465
Db 2116186 -----ACATACGGTGGTGAAGAGCGCTATATGATTTTAAATAATAATTA----- 2116142
QY 466 ThrArgIleHisSerThrLeuAspAsnGlnAsnMetVal-----lleThrLys 483
Db 2116141 ---CTGTAGCAGAACCACTAGATTGTAACATATATACGAGTTAACACATTTAAATTTA 2116085
QY 484 AlaArgGlnAsnLysTrpGluThrGly----- 492
Db 2116084 CGTAGTGAATCTAAATGGGACAGCAGCATAGCAAGCAAGTGTACCTGAAAGCGCTAAAGTA 2116025
QY 493 -----AsnAsnLeuAspValLeuThrTyrThrAsnProSer 504
Db 2116024 AAAGTTGAATGACCAAGATGATGTAATGGTACAAAGTAACTTATCAAAATAAACA 2115965
QY 505 -----TyrProAsnLeuAspHisGlnArgSerValLeuPheIleAsnLysTyrPhe 522
Db 2115964 GGTATATGCGGCTAATGATTTATTTATCTGAAAGTCTGCTGTGTAAAAAATTTAT 2115905
QY 523 LeuValIleAspArgAlaIleGlyGluAlaThrGlyAsnLeuGlyValHisTrpGlnLeu 542

Db 2115904 -----GCAARAGATAATTTAACTTACGTAGCGAAGCA 2115872
QY 543 LysGluAspSerAsnProValPheAspLysThrLysAsnArgValTyrThrTyrArg 562
Db 2115871 AAATGGGATAGT-----GAAATTTCTCAAGTAGTAGAA 2115839
QY 563 AspGlyAsnAsnLeuMetIleGlnSerLeuAsnAlaAspArgThrSerLeuAsn----- 580
Db 2115838 AAAGGTGAAAAGTAACTATTATTCG-----AAACAAGTATAAATGTTGG 2115791
QY 581 -----GluGluGluGlyLysValSerTyrValTyrAsnLysGluLeuLysArgPro 597
Db 2115790 CATGAAGTAACTTACGGTGGTAAAGAGGTATATG-----ATTTAAGTGAT 2115743
QY 598 AlaPheValPheGluLysPro 604
Db 2115742 AACTATTATAGTCGAAAACCA 2115722
RESULT 9
ABN70606
ID ABN70606 standard; DNA; 3495 BP.
XX
AC ABN70606;
XX
XX
DT 01-JUL-2002 (first entry)
XX
DE Streptococcus polynucleotide SEQ ID NO 9125.
KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
OS Streptococcus pyogenes.
XX
PN WO200234771-A2.
XX
PD 02-MAY-2002.
XX
PF 29-OCT-2001; 2001WO-GB04789.
XX
PR 27-OCT-2000; 2000GB-0026333.
PR 24-NOV-2000; 2000GB-0028727.
PR 07-MAR-2001; 2001GB-0005640.
XX
XX (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
XX Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
PI Tettelin H;
XX
XX WPI: 2002-352536/38.
DR P-PSDE; ABP29975.
XX
XX New Streptococcus protein for the treatment or prevention of infection
PT or disease caused by Streptococcus bacteria, such as meningitis, and
PT for detecting a compound that binds to the protein -
XX
XX Claim 7; Page 4034; 4525pp; English.
XX
CC The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic

CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.

XX Sequence 3495 BP; 1097 A; 733 C; 771 G; 894 T; 0 other;
 SQ

Alignment Scores:

Pred. No.: 0.0045 Length: 3495
 Score: 137.50 Matches: 152
 Percent Similarity: 33.05% Conservative: 85
 Best Local Similarity: 21.20% Mismatches: 263
 Query Match: 3.94% Indels: 217
 DB: 24 Gaps: 39

US-09-802-285A-2 (1-659) x ABN70606 (1-3495)

QY 43 GlyLeuGluLysVal-----AsnLysAlaValAlaAlaGlyAsn--- 55
 Db 1453 GGTCAAGAAAAGTCATGGTTTGGATCCTTACGCCAAATCTCTCGCTGCTGGAATGAT 1512
 QY 56 -----TyrAspAlaAlaLysAlaLeuLeuAlaTyrTyrArgGluLysSerLys--- 72
 Db 1513 GCGACTGCTACTGATGATCAATCAAAACAGCAAAAGCTGCTTTATTGATCCAAAGCAAACTA 1572
 QY 73 -----AlaArgGluProAspPheSerAsnAlaGluLysProAla 85
 Db 1573 GGACCAACAGCGCTTGATTTGGCAAAATTAACAACTTTAAAGCGTGAAGCGCTATT 1632
 QY 86 AspIleArgGlnProIleAspLysValThrArgGluMetAla---AspLysAlaLeuVal 104
 Db 1633 ATCTATGAAGCACATGCGCAGATTTTACGTCAGATAAGCTCTAGAAGCGCAAGTTAACA 1692
 QY 105 HisGlnPheGlnProHisLysGlyTyr---GlyTyrPheAspTyrGlyLysAspIle--- 122
 Db 1693 CACCCCTTTGGGACTTTTTCAGCTTTTGGTGAACAGCTAGACTATCTCAAAAGACTTGGGG 1752
 QY 123 -----AsnTrpGlnMetTrpProValLysAspAsnGluValArgTrpGlnLeuHisArg 140
 Db 1753 GTTACCACAGCTTCAATGTGTACCGTT----- 1779
 QY 141 ValLysTrpTrpGlnAlaMetAlaLeuValTyrHisAlaThrGlyAspGluLysTyrAla 160
 Db 1780 -----TTGAGTTATTTTATCCCAATGAGCTGGCAAGAGC 1815
 QY 161 ArgGluTrpValTyrGlnTyrSerAspTrpAlaArgLysAsnProLeuGlyLeuSerGln 180
 Db 1816 CGCTCAACAGCCTACAGCTCTTCAGAC----- 1842
 QY 181 AspAsnAspLysPheValTrpArgProLeuGluValSerAspArgValGlnSerLeuPro 200
 Db 1843 AATAATTAACAATGGGGTTATGACCA-----CAACACTATTTTGGCCCTTCT 1890
 QY 201 ProThrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGluPhe 220
 Db 1891 GGCATGTAATTCGGCAAAATCCTAATGACCTGCTTTACGTTATCGCAGAGCTTAAACCTT 1950
 QY 221 LeuAsnSerTyrHisGlnGlnAla-----AspTyrLeuSerThrHisTyr 235
 Db 1951 GTCAATGAGATTCAAAACGTGGTATGGGTGTTTATTTTGTGTGTTTATATACCAACAGC 2010
 QY 236 AlaGluGlnGlyAsnHisArgLeuPheGlu----- 245
 Db 2011 GCTAGAAC-----TATCTTTTGAAGATTGGAACCCCACTATATCAATTTTATG 2061
 QY 246 -----AlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPheLys 260
 Db 2062 AATGCTGATGCTACAGCTAGAGAGGTTTGGCGAGGTGCTAGGAACGACACATGCC 2121
 QY 261 AspSerProArgTrpArgGlnThrGlyIleSerValLeuAsnThrGluLeuLysGln 280
 Db 2122 ATGAGTCGTGATCTTTGGTGGATTCGATTACTTATCTGACTGCTGTAATCAAG----- 2175

QY 281 ValTyrAlaAspGlyMetGlnPheGluLeuSerProIleTyrHisValAlaAlaIleAsp 300
 Db 2176 -----GTAGATGGTTTTCGTTTCGATGATGGGTGACCATGATGGCGAGCTATTGAG 2229
 QY 301 IlePheLeuLysAlaTyrGlySerAlaLysArgValAsnLeuGluLysGluPheProGln 320
 Db 2230 -----CAAGCCTTAAGGCAGCCAAAGCCATTAA 2259
 QY 321 SerTyrValGlnThrValGluAsnMetIleMetAlaLeuIleSerLeuLeuProAsp 340
 Db 2259 ----- 2259
 QY 341 TyrAsnThrProMetPheGlyAspSerTrpIleThr-----AspLysAsnPhe 356
 Db 2260 CCAATACCATTAATGATTGGCAAGGCTGGCTACTCAAGGTGATGAGGGGAAAAA 2319
 QY 357 ArgMetAlaGlnPheAlaSerTrpAlaArgValPheProAlaAsnGlnAlaIleLysTyr 376
 Db 2320 GAAATTCGGCAGATCAAGATTGGATGAAA-----GCAACCAATACGTCGGTGT 2370
 QY 377 PheAlaThrAspGlyLysGlnGlyLysAlaProAsnPheLeuSerLysAlaLeuSerAsn 396
 Db 2371 TTCTCTGATGATATCAGA-----AATACCCTCAAGTCAGGTTTTTCCAAT 2415
 QY 397 AlaGlyPheTyrThrPheArgSerGlyTrpAspLysAsn-----AlaThrValMetVal 414
 Db 2416 GAAGGCACAGCAGCCTTTATTACTGTGGCGCAAAAAATCTAGAAGGTTTTTATCAAAACG 2475
 QY 415 LeuLysAlaSerPro-----ProGly-----GluPheHis 424
 Db 2476 ATCAAGCACAGCTGGTAACCTTTGAAGCAGATGCCCGAGGATGTAGTCAGATATT 2535
 QY 425 AlaGlnProAspAsnGlyThrPhe---GluLeuPheIleLysGlyArgAsnPheThrPro 443
 Db 2536 GCAGCCCATGACACCTGACCTTACATGATGTTGTCGCAATCCATCAATAAGGATCCT 2595
 QY 444 AspAlaGlyValPheValTyrSerGlyAspGluAlaIleMetLysLeuArgAsnTrpTyr 463
 Db 2596 AAAGTG-----GCTGAAGAAGAGATT-----CAC 2619
 QY 464 ArgGlnThrArgIleHisSerThrLeuThrLeuAspAsnGlnAsnMetValIleThrLys 483
 Db 2620 AAGCGTTTTCGTAGCAATACCATGATTTTAACTGCTCAAGGACTGCTTTATCCAT 2679
 QY 484 AlaArgGlnAsnLysTrpGluThrGlyAsnAsnLeuAspValLeuThrTyrThrAsnPro 503
 Db 2680 TCTGGTCAG-----GAATATGCAAGCAACCAAGCAGCTTCTA-----AATCCC 2721
 QY 504 SerTyr-----ProAsnLeuAspHisGlnArgSerVal 514
 Db 2722 GACTACAAGCAAAAGCGCTGATGACAAGGTGCCAAAT-----AAGGGACT 2769
 QY 515 LeuPhe-----IleAsnLysLysTyrPheLeuValIleAspArgAlaIleGlyGlu 531
 Db 2770 CTGATTGATGCTAGCGCAATACCTTACTTCAATCCAGATTCTTATGATTCGTCTGAT 2829
 QY 532 AlaThrGlyAsnLeuGlyValHisTrpGlnLeuLysGluAspSer-----AsnProVal 549
 Db 2830 GCGGTCAATCAATTT-----GACTGGGCAAGGCAACAGATTCCATAGTCCCGGATT 2883
 QY 550 PheAspLysThrLysAsnArgValTyrThr-----ThrTyrArgAspGlyAsn 565
 Db 2884 AGCAACCAACAAAA-----GCCATATACAGGGAGTAATTCGTTGCTGCTCAACA 2937
 QY 566 AsnLeuMetIleGlnSerLeuAsnAlaAsp----- 575
 Db 2938 GATGCTTTTACAAAGCAACCAAGCTAGGTAGATGCGGATGTGACCTTGTATCCCAA 2997
 QY 576 -----ArgThrSerLeuAsnGluGluGlyLysValSerTyr-----ValTyrAsn 591
 Db 2998 GCAGGACAGATGGTATTTCACAAAGGACCTCATCATGGTGTACCAACAGCTGGCATCA 3057
 QY 592 LysGluLeuLysArgProAlaPheValPheGluLysProLysLysAsnAlaGlyThrGln 611

Db 94 GTATTATCTTTGCTAGTCTGCTATTGAGATTAGGTATTATTTACAAATAGCACAAAGATCT 153
Qy 28 SerIleThrArg-----LysAspPheAspHisIleAsnLeuGluTyrSerGlyLeuGlu 45
Db 154 CATTCAACAATAATCAAAACGATGAACATAACT----- 192
Qy 46 LysValasnLysAlaValAlaAlaGlyAsn-----Tyr 56
Db 193 ---GTTATGATCAGTACCAAGAGCCGAATACTAGATAGAAATGGCAAGTACTAGTT 249
Qy 57 AspAspAlaAlaLysAlaLeuLeuAlaTyrTyrArgGluLysSerLysAlaArgGluPro 76
Db 250 GATAATGCTTCAAGATGCTATTACATACACTAGAAACCGTAAACATCAACAAGGAA 309
Qy 77 AspPheSerAsnAlaGluLysProAlaAsp---IleArgGlnProIleAspLysValThr 95
Db 310 ARGTTAAATACCTGTAAGAACTGCAGAGATTAAATTAATGATGATACAGATAAATTA 369
Qy 96 ArgGluMetAlaAspLysAlaLeuValHisGlnPheGlnProHisLysGlyTyrGlyTyr 115
Db 370 GAGAGA-----GATAAA----- 381
Qy 116 PheAspTyrGlyLysAspIleAsnTrpGlnMetTrpPro-----ValLysAspAsn 132
Db 382 -----AAGGATTTTGGATTCAAATGTATCCGTCATTTGCTAAAAAGTT-AAAT 428
Qy 133 GluValArgTrpGlnLeuHisArgValLysTrpTrpGln----- 145
Db 429 GAGAAAGA-----ACAATTAATGTTAGAGATGCGAGTATTTCAAGAGCAATTTGA 482
Qy 146 -----AlaMet 147
Db 483 TACCCAACTTAGAGATAAATAGGAAAAAAACAATTTAAAAACAGTTAACTAAAAAGATT 542
Qy 148 AlaLeuValTyrHisAlaThrGlyAspGluLysTyrAla----- 160
Db 543 GCAAGTTTGTAGCAATTTATCGGGAATGAACGTCGGTCAACTCTAGATCCTCAACAAT 602
Qy 161 -----ArgGlu----- 162
Db 603 TAAAAATGAGACGTAAAGGAGAAAGATATGACGCCGTATCAACACAGCTTCTTAATT 662
Qy 163 ---TrpValTyrGlnTyrSer-AspTrpAlaArgLysAsnProLeuGlyLeuSerGlnAs 181
Db 663 ACCTGGTGTAATACTACAATGATGGGATAGAAAAATACCCATACGGT-----GA 713
Qy 181 pAsnAspLysPheValTrpArgProLeuGluValSerAspArgValGlnSerLeuProPr 201
Db 714 TACTTTAAGAGGCATATTTTGA-----GATGTGCGACTTCGACTGAAGGTATACCTAA 767
Qy 201 othrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGluPheLe 221
Db 768 AGAATTAACT-----GAACAATATT 788
Qy 221 uAsnSerTyrHisGlnAlaAspTyrIleuSerThrHisTyrAlaGluGlnGlyAsnHis 241
Db 789 ATCAAAAGGTATTACCAAGATGATCGGTGCGTAAATCTTATCTTGAATATCAATACGA 848
Qy 241 sArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPheLysAs 261
Db 849 AGATGTTCTTAAGACGCAAGAAACAAATG-----AAATATACAC 890
Qy 261 pSerProArgTrpArgGlnThrGlyIleSerValLeuAsnThrGluIleLysLysGlnVa 281
Db 891 TGATAAACTCGAAGAGTAATAAGTTCAGAACTTAACTTCTCGGCTCAAGA----- 942
Qy 281 lTyrAlaAspGlyMetGlnPheGluLeuSerProIleTyrHisValAlaAlaIleAspIl 301
Db 943 ---GGTCATGATTTTACAATTAAT-----ATAGATAT 971
Qy 301 ePheLeuLysAlaTyrGlySerAlaLysArgValAsn-----LeuGluLysGluPhePr 319
Db 972 TGATTACAG-----AAAAAGTAGATCTTTATTAGAAAAACAAATTC 1016

Qy 319 oGlnSerTyrValGlnThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuPr 339
Db 1017 TAAATTACGTAGTCAAGGTGCTAAGGATATGACAAATCGTAAATGTTGTTCCAAATCC 1076
Qy 339 oAspTyrAsnThrProMetPheGlyAspSerTrpIleThrAspLysAsnPheArgMetAl 359
Db 1077 TAAAAATGGAGACATCTCGCTATTGCAGGAAGCAAAATGTATAGCAAGGTAAACTCAA 1136
Qy 359 aGlnPhe-----AlaSerTrpAlaArgValPheProAlaAsnGlnAlaIleLys---Ty 376
Db 1137 AGATTATGATATCGGCAACTTTTACAGCTCAATACACAGTAGGTCTTCTCAGTAAAGGAG 1196
Qy 376 rPheAlaThrAspGlyLysGlnGlyLysAlaProAsn-----PheLeuSerLys 392
Db 1197 AACATTATTAGCTGGATACCAAAATAAGCTATTATGTGTGAGAAACATCGTAGATGA 1256
Qy 392 sAlaLeuSerAsnAlaGlyPheTyrThrPheArgSerGlyTrpAspLysAsnAla----- 410
Db 1257 GCCATTAAAAATTCCAAGGTGGTTTAACTAAGCGTCTTATTATTAATAAAATGGTCATGT 1316
Qy 411 -----ThrVa 412
Db 1317 ATCTATCGATGATAAACAAGCACCTTATGCAITTCATCAAAACGTATACATGTTTAAACCCG 1376
Qy 412 lMetValLeuLysAlaSerProProGlyGluPheHisAlaGlnProAspAsnGlyThrPh 432
Db 1377 ACTTAAATTAGCAGGTGACCCCTATCTTACGTATGCTATTCACCTTAATAT----- 1428
Qy 432 eGluLeuPheIleLysGlyArgAsnPheThrProAspAlaGlyValPheValTyrSerGl 452
Db 1429 -----ATACGATGCTGCT----- 1443
Qy 452 yAspGluAlaIleMetLysLeuArgAsnTrpTyrArgGlnThrArgIleHisSerThrLe 472
Db 1444 -----CGTAAATTCGTAAGGGTTAAATCAAGTAGTCTTGGTGTAAAAAC 1490
Qy 472 uThrLeuAspAsnGlnAsnMetValIleThrLysAlaArgGlnAsnLysTrpGluThrGl 492
Db 1491 AGGTATTGACTTACCGAACGAAACCCAGGCCAAATAGAACCTTAACTAAATAATCCTGG 1550
Qy 492 yAsnAsnLeuAspValLeuThrTyrThrAsnProSerTyrProAsnLeuAspHisGlnAr 512
Db 1551 TAACTATTTAGACTTAGCTATTGGACAATACGACACATATACACCACTTCAGTTGCCCA 1610
Qy 512 gSerValLeuPheIleAsnLysLysTyrPheLeuVal-----IleAspArgAlaIl 529
Db 1611 ATACGTATCAACTATTGCTAATGATGGCTATAGAATTCAACACATATATGGATTGTCTAT 1670
Qy 529 eGlyGluAlaThrGlyAsnLeuGlyValHisTrpGlnLysGluAspSerAsnProVa 549
Db 1671 TTATGAATCTACTAAT-----AAAGATGAACCTGGTCCATT 1706
Qy 549 lPheAspLysThrLysAsnArgValTyrThrTyrArgAspGlyAsnAsnLeuMetIl 569
Db 1707 AAAACGTAATAATTAAGGGTAATGTTTAAATAAGGTAAATAACTCAATAGCAGAAATTA 1766
Qy 569 eGlnSerLeuAsnAlaAspArgThrSerLeuAsnGluGluGlyLysValSerTyrV 589
Db 1767 AGAAGTTCAGAAAGCTTCAAAATGCTTTCATCAAAAGCAAGGTACAGGTATGCTAG 1826
Qy 589 al-----TyrAsnLysGluLeuLysArgProAlaPheValPheGluLysP 604
Db 1827 TTTTAGAATACTAGTAGTACCTTCAGCTGGTGAACAGGAAGTCTGTAAGTTTTCAGA 1886
Qy 604 roLysLysAsnAlaGlyThrGlnAsnPheValSerIleValTyrPro 619
Db 1887 CGGAGAACCTAGAGTTAACTCAACAT---ATATCGTATTATGCACCCG 1930

RESULT 12

AAN40009

ID AAN40009 standard; DNA; 3994 BP.

XX


```
Db 3063 GAGTTTGATGTTG-----|||:::||||:::GTTTACGCTCTTAACACC 3092
Qy 453 AspGluAlaIleMetLys-----|||:::||||:::LeuArgAsnTrpTyrArgGlnThrArg 467
Db 3093 ACACAAACCATATCCAAATCCGCTCTTTGCCGACTTATCATCTTGGTTCAAGGGTTTGAA 3152
Qy 468 -----|||:::||||:::IleHisSerThrLeuThrLeuAsp 475
Db 3153 GATCCTGAAGATATTTGAGATGGGTTTGAAGTCAGTCTCTCTCTCTCTCTCTCTTTTGGAC 3212
Qy 476 AsnGlnAsnMetValIleThrLysAlaArgGlnAsnLysTrpGluThrGly-----|||:::||||::: 492
Db 3213 CGTGTACTCTAAGTCAAGTTCTCAGGAGACCCATATTTCCACAAACAGATGCT 3272
Qy 493 ---AsnAsnLeuAspValLeuThrTyrThrAsnProSerTyrProAsn-----|||:::||||::: 507
Db 3273 GTCACAAACCAACCATTCAGTCTCAGAACGACCTAAGTTACTATAAAGTGTACGGCCTA 3332
Qy 508 LeuAspHisGlnArgSerValLeuPheIleAsn-----|||:::||||:::LysLys 520
Db 3333 CTGGATCAAAACATCTTGAATGTACTTCAACGATGGAGATGCTGTTTCTACAATACC 3392
Qy 521 TyrPheLeuValIleAspArgAlaIleGlyGluAlaThrGlyAsnLeuGlyValHisTrp 540
Db 3393 TACTTCATGACACCCGTAACGCTCTAGATCTGTGAACATGACCACTGGTGTCTC-----|||:::||||::: 3446
Qy 541 GlnLeuLysGluAspSerAsnProValPheAspLys-ThrLys-----|||:::||||:::AsnArg 556
Db 3447 -----GATAATTTGTTTCTACATTCACAAGTTCCAAAGTTCCAAAGTAAGGAAGTAAATAATAG 3494
Qy 556 gValTyrThrThrTyr 561
Db 3495 AGTTATAAACTTAT 3510

RESULT 13
ID AAQ11883 standard; DNA; 3994 BP.
XX AAQ11883;
AC AAQ11883;
XX 24-JUL-1991 (first entry)
XX Invertase gene.
XX Invertase; yeast; promoter; expression vehicle; interferon; ss.
XX Saccharomyces cerevisiae.
XX Key Location/Qualifiers
FT CDS 1896..3491
FT /*tag= a
FT /product= invertase
XX
XX US010003-A.
XX
XX 23-APR-1991.
XX
XX 20-JUN-1990; 90US-0541186.
XX
XX 25-APR-1983; 83US-0488387.
XX
XX 20-JUN-1990; 90US-0541184.
XX
XX (GETH ) GENENTECH INC.
XX
XX Chang CN, Hitzeman RA, Matteucci MD;
XX
XX WPI; 1991-148222/20.
XX
XX P-PSDB; AAR12145.
XX
XX Yeast expression vehicle - contg. homologous yeast signal peptide
XX for processing and export of mature heterologous protein from
XX yeast cells
```

```
XX Disclosure; Fig 13 (A-C); 29pp; English.
PS
XX
CC The invertase promoter and signal sequence is used in the
CC construction of a yeast expression vehicle which also comprises a DNA
CC sequence encoding a protein heterologous to the yeast organism. The
CC homologous yeast signal peptide is sufficient, upon expression as a
CC fusion protein contg, the signal peptide and heterologous protein, to
CC result in the processing and export of mature heterologous protein into
CC the culture medium of yeast organisms transformed with the expression
CC vehicle, the heterologous protein being completely processed to the
CC mature form.
CC See also AAQ11885-87.
XX
SQ Sequence 3994 BP; 1201 A; 773 C; 768 G; 1252 T; 0 other;

Alignment Scores:
Pred. No.: 0.0573 Length: 3994
Score: 127.00 Matches: 118
Percent Similarity: 32.34% Conservative: 78
Best Local Similarity: 19.47% Mismatches: 185
Query Match: 3.63% Indels: 225
DB: 12 Gaps: 33

US-09-802-285A-2 (1-659) x AAQ11883 (1-3994)
Qy 94 ValThrArgGluMetAlaAspLysAlaLeuValHisGlnPheGlnProHisGlyTyr 113
Db 1956 ATGACAAACGAACTAGCGATAGACTTTGTGTCCAC---TTACACCCCAACAGGCTGG 2012
Qy 114 -----GlyTyrPheAspTyrGlyLysAspIleAsnTrpGlnMetTrp--- 127
Db 2013 ATGATGACCCCAATGGTGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 2072
Qy 128 ProValLysAspAsnGluValArgTrpGlnLeuHisArgValLysTrpGlnAlaMet 147
Db 2073 CAATACAAACCAATGACACCGTATGG-----GGTACGCCATTTGTTTGGGCGC--- 2120
Qy 148 AlaLeuValTyrHisAlaThrGlyAspGluLysTyrAlaArgGluTyrValTyrGlnTyr 167
Db 2121 -----CATGCTACTTCCGATGAT-----TTG 2141
Qy 168 SerAspTrpAlaArgLysAsnProLeuGlyLeuSerGlnAspAsnAspLysPheValTrp 187
Db 2142 ACTAATTCG---GAAGATCAACCCATTCGTATCGTCCCAACGCGTAAC----- 2186
Qy 188 ArgProLeuGluValSerAspArgValGlnSerLeuProThrPheSerLeuPheVal 207
Db 2186 ----- 2186
Qy 208 AsnSerProAlaPheThrProAlaPheLeuMetGluPhe----- 220
Db 2187 GATTCAGGTGCTTCTCTGCGTCCCATGTTGTTGATTACAAACACGAGTGGGTTTTC 2246
Qy 221 -----LeuAsnSerTyrHis 225
Db 2247 AATGATACTATTGATCCAAAGACAAAGATGCGTTCGATTTGGACTTATAACACTCTGAA 2306
Qy 226 GlnGlnAlaAspTyrLeuSerThrHisTyrAlaGluGlnGlyAsnHisArgLeuPheGlu 245
Db 2307 AGTGAAGAGCAATACATTAGC-----TATTCTCTGATGGTGGTTACACTTTTACTGAA 2360
Qy 246 AlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPheLysAsp----- 261
Db 2361 TACCAAAAGACCCCTGTTTAGTCCCAACTCCACTCAATTCAGAGATCCAAAGGTGTTTC 2420
Qy 262 -----SerProArgTrpArgGlnThrGlyIleSerValLeuAsnThrGluIle 277
Db 2421 TGGTATGAACCTCTCTCAAAAATGGATTATGACG-----GCTGCCAAATCACAAGACTAC 2474
Qy 278 LysLysGlnValTyrAlaAsp----- 284
Db 2475 AAAATTGAAATTTACTCTCTGATGACTTGAAGTCTCTGGAAGCTAGATCTGCATTGGCC 2534
```

```

QY 285 -----GlyMetGlnPheGluLeu----- 290
Db 2535 AATGAAGTTTCTTAGCTACCAATACGAATGTCAGGTTGATGAAGTCCCAACTGAG 2594
QY 291 -----SerProIleTyrHisValAlaAlaIleAspIle----- 301
Db 2595 CAAGATCCTTCCAAATCTTATGGTCTATGTTTATTTCTATCAACCCAGGTGCACCTGCT 2654
QY 302 -----PheLeuLysAlaTyr 306
Db 2655 GGCAGTTCTTCAACCAATATTTTGTGATCCTCAATGCTACTCATTTTGAAGCGTTT 2714
QY 307 GlySerAlaLysArgValAsnLeuGluLysGluPheProGlnSerTyr-----ValGln 324
Db 2715 GACATCAATCTAGAGTG-----GTAGATTTTGGTAGAGCTACTATGCTGCTTCGAA 2765
QY 325 ThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuProAspTyrAsnThrPro 344
Db 2766 ACTTCTCTCAAC-----ACTGACCCCAACTACGGTTCAGCA 2801
QY 345 MetPheGlyAspSerTrpIleThrAspLysAsnPheArgMetAlaGlnPheAla----- 362
Db 2802 TTA---GGTATTCCTGGGCTTCA-----AACTGGGAGTACAGTGCCTTTGTCCCAACT 2852
QY 363 -----SerTrpAlaArgValPheProAlaAsnGlnAlaIleLys 375
Db 2853 AACCCATGGAGATCATCCTGCTTTGGTCCGCAAGTTTCTTTGAAC-----ACTGAA 2906
QY 376 TyrPheAla-----ThrAspGlyLysGlnGlyLysAlaProAsnPheLeuSerLys 392
Db 2907 TATCAAGCTAATCCAGAGACTGAATGATCAATTTGAAGCCGCAACCAATATTGAAC--- 2963
QY 393 AlaLeuSerAsnAlaGlyPheTyrThrPheArgSerGlyTrpAspLysAsnAlaThrVal 412
Db 2964 ---ATTAGTAATGCTGGT-----CCCTGGTCTCGTTTGTGTTACTTAAC 3002
QY 413 MetValLeuLysAlaSerProProGlyGluPheHisAlaGlnProAsnPheThrPhe 432
Db 3003 ACAACTCTAAGGCAATCTTACATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGAT 3062
QY 433 GluLeuPheIleLysGlyArgAsnPheThrProAspAlaGlyValPheValTyrSerGly 452
Db 3063 GAGTTTGAGTTG-----GTTTACGCTGTATACACC 3092
QY 453 AspGluAlaIleMetLys-----LeuArgAsnTrpTyrArgGlnThrArg 467
Db 3093 ACACAAACCAATATCCAAATCCGCTTTTGGCGACTTATCACTTTTGGTCAAGGGTTAGAA 3152
QY 468 -----IleHisSerThrLeuThrLeuAsp 475
Db 3153 GATCCTGAGAAATATTGAGAATGGGTTTGAAGTCAGTCTCTCTCTCTCTCTCTCTCTCTCT 3212
QY 476 AsnGlnAsnMetValIleThrLysAlaArgGlnAsnLysTrpGluThrGly----- 492
Db 3213 CGTGGTAATCTTAAGTCAAGTTTGTCAAGGAGAACCCATATTTCACAAACAGAAATGCT 3272
QY 493 ---AsnAsnLeuAspValLeuThrTyrThrAsnProSerTyrTrpProAsn----- 507
Db 3273 GTCACACCAACCATTCAGTCTGAGAACGACCTAAGTTACTATATAAGTGTACGGCCTA 3332
QY 508 LeuAspHisGlnArgSerValLeuPheIleAsn-----LysLys 520
Db 3333 CTGGATCAAAACATCTTGGNAATTTGACTTCAACGATGGAGATGTTGTTCTTACAAATACC 3392
QY 521 TyrPheLeuValIleAspArgAlaIleGlyGluAlaThrGlyAsnLeuGlyValHisTrp 540
Db 3393 TACTTCATGACCAACCGCTAAGCTCTAGGATCTGTGAACATGACCACTGGTGTCT----- 3446
QY 541 GlnLeuLysGluAspSerAsnProValPheAspLys-ThrLys-----AsnAr 556
Db 3447 -----GATAATTGTTTCTACATTGACAGATTCCAGTAAGGGAAGTAATAATAG 3494

```

```

QY 556 gValTyrThrThrTyr 561
Db 3495 AGGTATAAAACTTAT 3510
RESULT 14
AAQ41744
ID AAQ41744 standard; DNA; 3414 BP.
XX AAQ41744;
XX
XX 14-SEP-1993 (first entry)
XX
XX DNA encoding B.t. toxin HD511.
XX
XX Coleopteran pests; Bacillus thuringiensis; insecticide; transgenic;
XX plant; pesticide; Colorado potato beetle; ss.
XX
XX Bacillus thuringiensis strain HD511.
XX
XX WO9308693-A.
XX
XX 13-MAY-1993.
XX
XX 06-NOV-1992; 92WO-US09510.
XX
XX 06-NOV-1991; 91US-0788638.
XX
XX (MYCO ) MYCOGEN CORP.
XX
XX Fu JM, Payne JM;
XX
XX WPI; 1993-167285/20.
XX
XX P-PSDB; AAR37213.
XX
XX Control of coleopteran insect pests using Bacillus thuringiensis
XX - using strains Bt HD 511, HD 867 or HD 1011 or their toxins or
XX nucleotide sequences
XX
XX Claim 13; Page 13-15; 29pp; English.
XX
XX The DNA sequence encoding Bacillus thuringiensis strain HD 511 toxin
XX may be used in an insecticide/pesticide for treatment of Coleopteran
XX pests, esp. the Colorado potato beetle. Strain HD511 and recombinant
XX cells contg. the gene encoding the toxin may be treated by chemical or
XX physical means to prolong the pesticidal activity of the cells, the
XX treated cells acting as a protective coating for the pesticidal toxin,
XX which becomes available upon ingestion by a target insect. The compsn.
XX may be applied to the environment of the coleoptera, e.g. plants, soil
XX or water, by spraying, dusting, etc.
XX See also AAQ41745.
XX
SQ Sequence 3414 BP; 1215 A; 534 C; 658 G; 1007 T; 0 other;

Alignment Scores:
Pred. No.: 0.101 Length: 3414
Score: 123.50 Matches: 141
Percent Similarity: 35.92% Conservative: 95
Best Local Similarity: 21.46% Mismatches: 199
Query Match: 3.53% Indels: 224
DB: 14 Gaps: 39

US-09-802-285A-2 (1-659) x AAQ41744 (1-3414)
QY 48 AsnLysAlaValAla-----AlaGlyAsnTyrAspAlaAlaLysAlaLeu 63
Db 361 AATAAGCACTTGCAGAAATTAGAGGATTAGGAATAAC----- 399
QY 64 LeuAlaTyrTyrArgGluLysSerLysAlaArgGluProAspPheSerAsnAlaGluLys 83
Db 400 TTAACAATATATCAACAG-----GCACCTTGAAGATTGGCTGAACAATCTGTATGAT 450
QY 84 ProAlaAspIleArgGlnProIleAspLysValThrArgGluMetAlaAspLysAlaLeu 103

```

```

451 CCAGCACTATAACACGAGTGATAGATCGTTT-----CGTATTATAGAT---GCTTTA 501
QY
104 ValHisGlnPheGlnProHisLysGlyTyrGlyTyrPheAsp----- 117
Db
502 TTGGAATCATATATATCCCGTCATTATAGGGTGTGGATATGAATAATCACTAACAGTT 561
QY
118 TyrGlyLysAspIleAsnTrpGlnMet----- 126
Db
562 TACGCACAAGCGGCAACCTTCATCTAGCTTTATTAAAGAGATTCTACTCTTTATGGAGAT 621
QY
127 ---TrpProValLysAspAsnGluVal----- 134
Db
622 MAATGGGATTCACTCAGAACACATTTAGGAAAATTATAATCGTCAAAAGAAACATATC 681
QY
135 ---ArgTrpGlnLeuHisArgValLysTrpTrpGlnAla---MetAlaLeuValTyrHis 152
Db
682 TCTGAATATCTAACACATTCGCTTAAGTGGTATAATAGTGGCTTTAGCGAGATTGAACGGT 741
QY
153 AlaThrGlyAspGlu-----LysTyrAlaArgGluTrpValTyrGlnTyr 167
Db
742 TCCACTTATGAACATGGTAATAATTATATATCGTTTCGTAGAGAAATGATATTAATGGTA 801
QY
168 SerAspTrpAlaArgLysAsnProLeu-----GlyLeu 178
Db
802 TTAGATATTGCTGTGTTATTCCTATTATGACCCCTCGAATGTTATTCATGGAACAAGT 861
QY
179 SerGlnAspAsnAspLysPheValTrpArg-ProLeuGluValSerAspArgValGlnSe 198
Db
862 ACGCAGTT-AACGAGAGAAGTGTATACCGATCCATTAGCTTGTCAATTAGCAATCCAGA 920
QY
198 rleuProThrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMe 218
Db
921 TATAGTCCAGCTTTTCTCAGATGAAATAACTGCGTTTAGAACACACACACTTGT-- 978
QY
218 tgluPheLeuAsnSerTyrHisGlnAlaAlaAspTyrLeuSerThrHisTyrAlaGlu 238
Db
979 -----GATTATTATAGATGAGCTTTATATATATATAC 1007
QY
238 nGlyAsnHisArgLeuPhe-----GluAlaGlnArgAsnLeuPhe-----Al 252
Db
1008 ATCAAAATAAAGCATTTTTCATGAGATTCAACGACCATTTATTTATGGTGTGTACA 1067
QY
252 aGlyValSerPheProGluPheLysAspSerProArgTrpArgGlnThrGlyLe-SerV 272
Db
1068 TAAGGTAGCTTTAAATAATCGGACCAATCCAATTATAT--ACAACAGGCATATATG 1124
QY
272 alLeuAsnThrGluLe-----LysLysGlnValTyrAlaAspGlyMet-----GlnPheG 289
Db
1125 TAAACAGTGGATATATTTCATCAGAGCATATTTCATTTAGAGGGAATGATATCTATAG 1184
QY
289 luleuSerProIleTyrHisValAlaAlaIleAspIlePheLeuLysAlaTyrGlySerA 309
Db
1185 ACATTAGCAGCTCCATCATGTTGTAGTTTATCCGTATA--CTCAGANTATGGTGTGCA 1241
QY
309 lalys----- 310
Db
1242 GCAAGTTGAGTTTACGGGTAAAGGCGCATGPACATTATAGAGGAGATAACAATAATGA 1301
QY
311 -----ArgValAsnLeuGluLysGluPheProGlnSerTyrValGlnThrValGluAsnM 329
Db
1302 TCTGAGTATGATTCTTATTGATCAATTACCCCGAGGAGACCAACCAACAGAAAATA 1361
QY
329 etileMetAlaLeu-----IleSerIleSerLeuProAspTyrAsnThr- 343
Db
1362 CACTCATCGA-TTATGTCATGTCAGCTATATCTAAATCACTCCGGATTATGATAATG 1420
QY
344 -----ProMetPheGlyAspSerTrpIleThrAspLysAsnPheArgMetAlaGlnP 361
Db
1421 CTACTATCCGATCTTT-----TCTTGG-----ACGCATAGAAGTGGCGAGT 1462
QY
361 heAlaSerTrpAlaArgValPheProAlaAsn-----GlnAlaIleLysTyrP 377
Db
1463 AT-----TACAATAGATATCTATCCAAACAAATAATCAAAAAATTCACGCTGTAAAAATGT 1516

```

```

377 heAlaThrAspGlyLysGlnGly-----LysAlaProAsnPheLeuSerLysAlaLeuS 395
QY
1517 ATAAACTAGATCACTACTACTACAGTTGTCAAAAGGCGCTGGATTACAGGTGAGATTAG 1576
Db
395 erAsnAlaGlyPheTyrThrPheArgSerGlyTyr-----AspLysAsnAlaThrValM 413
QY
1577 TTAAGAGAGGG-----AGTAATGGTTATATAGGAGATATAAAGGCTACCGTA- 1623
Db
413 etValLeuLysAlaSerProProGlyGlu-----PheHisAlaGlnP 427
QY
1624 -----AACCTCACCATCTTCTCAAAATATCGTTAGAGTTTCATACGCCACTA 1672
Db
427 toAspAsnGlyThrPheGluLeuPheIleLysGlyArgAsnPheThrProAspAlaGlyV 447
QY
1673 GTGTTTCTGGACTATTCAACGTTTAT----- 1701
Db
447 alPheValTyrSerGlyAspGluAlaIleMetLysLeuArgAsnTrpTyrArgGlnThrA 467
QY
1701 ----- 1701
Db
467 rgIleHisSerThrLeuThrLeuAspAsnGlnAsnMetValIleThrLysAlaArgGlnA 487
QY
1702 -----AATGATGAATAGCGCTTCAAAAAAATTTTCAAA 1735
Db
487 snLysTrpGluThr---GlyAsnAsnLeuAspValLeuThrTyr----- 500
QY
1736 GTACTGTAGAACATAATAGTGAAGGAAGAT---TTAACTATGTTTCATTTGGATATA 1792
Db
501 -----ThrAsnProSerTyrProAsn-----LeuA 509
QY
1793 TAGAATATTCTACGACCATTCATTTCCGAATGAGCATCCAAAAATCACTTCTCATTTAA 1852
Db
509 spHis-----GlnArgSerValLeuPheIleAsnLysTyrPheLeuValIleAspA 527
QY
1853 ACCATTGAGTAACAAATTCACCATTTTATGTAGATTCATCGAATTTATCCCTGTAGAT- 1911
Db
527 rgAlaIleGlyGluAlaThrGlyAsnLeuGlyValHisTrpGlnLeuLysGluAspSerA 547
QY
1912 -----GTAAATTATGATGAAAAAGAAAAA--- 1935
Db
547 snProValPheAspLysThrLysAsnArgValTyrThrTyrArgAspGlyAsnAsnL 567
QY
1936 -----CTAGAAAAAGCACAGAAAGCGGTGAATACCTTGTTCAGAGGGAAGAATG 1987
Db
567 euMet-----IleGlnSerLeuAsnAlaAspArgThrSerLeu 579
QY
1988 CACTCCAAAAATACGTGACAGATTATAAGTCGACCGGTTTCAATT 2034
Db

```

RESULT 15
AAV09317
ID AAV09317 standard; cDNA; 1542 BP.
XX AAV09317;
AC AAV09317;
XX
XX
DT 13-MAY-1998 (first entry)
DE Unsecreted yeast invertase gene (SUC2) used as a selection marker.
XX Secreted protein; unsecreted invertase; yeast; selection marker;
KW Isolation; ss.
XX Saccharomyces cerevisiae.
OS
XX
XX
XX US5712116-A.
PN
PD 27-JAN-1998.
XX
XX 24-APR-1996; 96US-0639052.
XX
XX 07-APR-1993; 93US-0045267.
PR 24-OCT-1994; 94US-0328962.
PR 24-APR-1996; 96US-0639052.

XX PA (GEM) GENETICS INST INC.

XX PI Jacobs K;

XX DR WPI; 1998-119980/11.

XX PT Isolation of cDNA encoding secreted protein - using DNA encoding

XX PT non-secreted yeast invertase as selection marker

XX PS Disclosure; Columns 5-8; 5pp; English.

XX CC This is the nucleotide sequence of an unsecreted yeast invertase from

XX CC the clone SUC2 of *Saccharomyces cerevisiae*. This is used in a method

XX CC for isolating a cDNA that encodes a secreted protein. The method

XX CC comprises constructing a cDNA library from cellular mRNA and ligating

XX CC the cDNA library to DNA encoding a non-secreted yeast invertase from

XX CC which the secretory leader sequence and initiating methionine have been

XX CC deleted. The ligation mixture is used to transform a yeast cell that

XX CC does not contain an invertase gene. Transformed yeast cells capable of

XX CC growing on sucrose or raffinose are selected and the DNA purified from

XX CC the selected yeast cells is sequenced. A second cDNA library is prepared

XX CC from cellular mRNA and is screened to detect full-length cDNA which is

XX CC then isolated. Novel secreted proteins (e.g. cytokines) can be

XX CC discovered using this method that requires neither a bioassay nor

XX CC knowledge of homology with other proteins.

XX SQ Sequence 1542 BP; 448 A; 337 C; 306 G; 451 T; 0 other;

Alignment Scores:

Pred. No.: 0.0382 Length: 1542

Score: 123.00 Matches: 109

Percent Similarity: 32.01% Conservative: 76

Best Local Similarity: 18.86% Mismatches: 179

Query Match: 3.52% Indels: 214

DB: 19 Gaps: 31

US-09-802-285A-2 (1-659) x AAV09317 (1-1542)

QY 94 ValThrArgGluMetAlaAspLysAlaLeuValHisGlnPheGlnProHisLysGlyTyr 113

DB 4 ATGACAAACGAACTAGCGATACACCTTTGGTCCAC---TTACACCCCAACAAAGGGCTGG 60

QY 114 -----GlyTyrPheAspTyrGlyLysAspIleAsnTyrGlnMetTyr-- 127

DB 61 ATGAATCACCACCAATGGTGTGGTAGCATGAAAGATGCCAAATGGCATCTGTACTTT 120

QY 128 ProValLysAspAsnGluValArgTyrGlnLeuHisArgValLysTyrTrpGlnAlaMet 147

DB 121 CAATACAAACCAATGACACCGGTATGG-----GGTAGCGCATTTGTTTGGGCG- 168

QY 148 AlaLeuValTyrHisAlaThrGlyAspGluLysTyrAlaArgGluTrpValTyrGlnTyr 167

DB 169 -----CATGCTACTTCGGATGAT-----TTG 189

QY 168 SerAspTyrAlaArgLysAsnProLeuGlyLeuSerGlnAspAsnAspLysPheValTyr 187

DB 190 ACTAATTTGG---GAAGATCAACCCATTGCTATCGCTCCCAAGCGTAAC----- 234

QY 188 ArgProLeuGluValSerAspArgValGlnSerLeuProProThrPheSerLeuPheVal 207

DB 234 ----- 234

QY 208 AsnSerProAlaPheThrProAlaPheLeuMetGluPhe----- 220

DB 235 GATTCAGGCTTCTCTCGCTCCATGGTGGTGTGATTACAAACACACAGATGGGTTTTC 294

QY 221 -----LeuAsnSerTyrHis 225

DB 295 AATGATACTATTGATCCACACAAAGATGCGTTGGCATTTGGACTTATACACTCTCTGAA 354

QY 226 GlnGlnAlaAspTyrLeuSerThrHisTyrAlaGluGlnGlyAsnHisArgLeuPheGlu 245

DB 226 GlnGlnAlaAspTyrLeuSerThrHisTyrAlaGluGlnGlyAsnHisArgLeuPheGlu 245

DB 355 AGTGAAGAGCATACATTAGC-----TATTCTCTTGATGGTGTACACTTTTACTGAA 408

QY 246 AlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPheLysAsp----- 261

DB 409 TACCAAAAGAACCTCTGTTTAGTCTGCCAATCCCACTCAATTCAGAGATCCAAAGGTGTTTC 468

QY 262 -----SerProArgTyrArgGlnThrGlyLeuSerValLeuAsnThrGluIle 277

DB 469 TGGTATGAACCTCTCTCAAAATGGATTATGACG-----GCTGCCAAATCACAGACTAC 522

QY 278 LysLysGlnValTyrAlaAsp----- 284

DB 523 AATATTGAATTTACTCTCTGATGACTTGGAGCTCTGGAGCTAGATACTGCATTTGCC 582

QY 285 -----GlyMetGlnPheGluLeu----- 290

DB 583 AACGAAGTTTCTTAGGCTACCAATACGAATGTCCAGGTTTGAATGAAGTCCCAACTGAG 642

QY 291 -----SerProIleTyrHisValAlaIleAspIle----- 301

DB 643 CAAGATCCTTCCAAATCTTATGGTCTATGTTTATTTCTATCACCCAGGTGCACCTGCT 702

QY 302 -----PheLeuLysAlaTyr 306

DB 703 GCGGTTCTTCAACCAATATTTTGGTGGATCTCTCAATGCTACTCATTTTGAAGCGTTT 762

QY 307 GlySerAlaLysArgValAsnLeuGluLysPheProGlnSerTyr-----ValGln 324

DB 763 GACAATCAATCTAGATG-----GTAGATTTTGGTAAGGACTACTATGCTTGCCTGCAA 813

QY 325 ThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuProAspTyrAsnThrPro 344

DB 814 ACTTCTTCAAC-----ACTGACCCACACTACGTTGTCAGCA 849

QY 345 MetPheGlyAspSerTyrPheThrAspLysAsnPheArgMetAlaGlnPheAla----- 362

DB 850 TTA---GGTATTCCTGGGCTTCA-----AATGGAGTACAGTGCCTTTGTCCCAACT 900

QY 363 -----SerTyrAlaArgValPheProAlaAsnGlnAlaIleLys 375

DB 901 AACCCATGGAGATCATCCATGCTTTGGTCCGCAAGTTTCTTTGTAAC-----ACTGAA 954

QY 376 TyrPheAla-----ThrAspGlyLysGlnGlyLysAlaProAsnPheLeuSerLys 392

DB 955 TATCAAGCTAATCCAGAGACTGAATGATCAATTTGAAGCCGCAACCAATATTTGAAC- 1011

QY 393 AlaLeuSerAsnAlaGlyPheTyrThrPheArgSerGlyTyrAspLysAsnAlaThrVal 412

DB 1012 ---ATTAGTAATGCTGCT-----CCCTGGTCTGTTTGTCTACTAAC 1050

QY 413 MetValLeuLysAlaSerProGlyGluPheHisAlaGlnProAspAsnGlyThrPhe 432

DB 1051 ACAACTCTAATAAGCCCAATCTTCAATGTCGATTGAGCAACTCGACTGTGACCTA 1110

QY 433 GluLeuPheIleLysGlyArgAsnPheThrProAspAlaGlyValPheValTyrSerGly 452

DB 1111 GAGTTTGGTTG-----GTTTACGCTGTAAACACC 1140

QY 453 AspGluAlaIleMetLys-----LeuArgAsnTyrTyrArgGlnThrArg 467

DB 1141 ACACAACCATATCCAAATCCGCTTTGCGGACTATCATCTTTGGTTCAAGGGTTTAGAA 1200

QY 468 -----IleHisSerThrLeuThrLeuAsp 475

DB 1201 GATCTGGAAGATAATTGAGAATGGTTTGAAGTCAGTCTCTCTCTCTTTTGGAC 1260

QY 476 AsnGlnAsnMetValIleThrLysAlaArgGlnAsnLysTyrGluThrGly----- 492

DB 1261 CGTGGTAACTTAAGGTCAAGTTTGTCAAGGAGAACCCATATTTCAACAACAGAAATGCT 1320

QY 493 ---AsnAsnLeuAspValLeuThrTyrThrAsnProSerTyrProAsn----- 507

DB 1321 GTCAACAACCAACCATTCAGTCTGAGAACGACCTAAGTTACTATAAAGTTACGCGCTA 1380

Qy 508 LeuaspHisGlnArgSerValIleuPheIleasn-----LysLys 520
 Db 1381 CTGGATCAAAACATCTTGGATTGTACTTCAACGATGGAGATGGTTTCTACAAATACC 1440
 Qy 521 TyrPheLeuValIleaspArgAlaIleGlyGluAlaThrGlyAsnLeuGlyVal 538
 Db 1441 TACTTCATGACCACCGGTAACGCTCTAGGATCTGTCAACATGACCACCTGGTCTC 1494

Search completed: August 4, 2003, 11:30:38
 Job time : 4005 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 4, 2003, 07:43:34 ; Search time 4666 Seconds
(without alignments)
4110.319 Million cell updates/sec

Title: US-09-802-285A-2
Perfect score: 3494
Sequence: 1 MTKIFKRIIFAVIALSSG.....KGLNLTTLNGKQQLVLVP 659

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Delop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlh

-MODEL=frame+p2n.model -DB=GenEmbl -QFMT=fastap -SURFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=PCT -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09802285 @CN 1 1 3052 @runat 28072003 190510.9705 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.or.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	3494	100.0	1980	1	CHU27586	U27586 Cytophaga h
2	3494	100.0	1980	6	I71365	I71365 Sequence 3
C 3	235.5	6.7	50327	1	AE014145	AE014145 Streptoco
C 4	234.5	6.7	12195	1	AE006517	AE006517 Streptoco
C 5	233.5	6.7	12187	1	AE010002	AE010002 Streptoco
6	198	5.7	175936	2	SPNEU1908	AL449930 Streptoco
7	189	5.4	11333	1	AE008410	AE008410 Streptoco
8	186.5	5.3	34496	1	SCF62	AL121855 Streptomy
9	177.5	5.1	22913	1	AF498407	AF498407 Pseudomon
C 10	176	5.0	3763	6	BD003866	BD003866 Polynucle
11	176	5.0	14385	1	AE007344	AE007344 Streptoco
12	176	5.0	19295	1	AF498414	AF498414 Pseudomon
13	170.5	4.9	16720	1	AF498401	AF498401 Pseudomon
14	170.5	4.9	16720	1	AF498411	AF498411 Pseudomon
C 15	160.5	4.6	205050	1	AL646082	AL646082 Ralstonia
16	154	4.4	2882	1	AE003330	AE003330 Pseudomon
C 17	145.5	4.2	10566	1	AE010107	AE010107 Streptoco
C 18	140.5	4.0	50354	1	AE014169	AE014169 Streptoco
19	139	4.0	2295	6	AX415476	AX415476 Sequence
C 20	139	4.0	23292	1	SCE134	AL049661 Streptomy
C 21	139	4.0	258650	1	AL596171	AL596171 Listeria
C 22	139	4.0	349980	6	AX417046	AX417046 Sequence
C 23	139	4.0	349980	6	AX417047	AX417047 Sequence
C 24	137.5	3.9	10586	1	AE006620	AE006620 Streptoco
C 25	137.5	3.9	26953	1	ALW243431	AJ243431 Acinetoba
C 26	134	3.8	321250	1	MPULM02	AL445564 Mycoplasma
27	132.5	3.8	67970	3	PFMAL1P3	AL031746 Plasmodiu
28	129.5	3.7	39305	8	SPU33009	U33009 Schizosacch
29	129.5	3.7	41799	8	SPFC1683	AL355920 Schizosac
C 30	129.5	3.7	85837	8	SPU33010	U33010 Schizosacch
31	129	3.7	9130	12	AF138275	AF138275 Synthetic
32	129	3.7	9193	12	AF138273	AF138273 Synthetic
33	129	3.7	9194	12	AF138274	AF138274 Synthetic
C 34	128	3.7	13280	1	AE013673	AE013673 Yersinia
35	128	3.7	148753	2	AC122951	AC122951 Rattus no
36	128	3.7	216050	1	AJ414157	AJ414157 Yersinia
37	127	3.6	1801	8	AV128845	AV128845 Arabidops
38	127	3.6	1993	8	AY081353	AY081353 Arabidops
39	127	3.6	2873	8	SCINVE	VO1311 Yeast gene
40	127	3.6	3994	6	E00284	E00284 DNA coding
41	126	3.6	3749	1	AF305610	AF305610 Borrelia
42	125	3.6	3072	1	AF367012	AF367012 Streptoco
43	124	3.5	4410	1	LEUORF15X	L28806 Leuconostoc
44	124	3.5	4410	1	LOPLOGFG	Z29976 L.oenosp pla
45	123.5	3.5	3414	1	BTU04367	U04367 Bacillus th

ALIGNMENTS

RESULT 1


```

Db      1201  |||||TAGAAGCGGATGGGATATAAAATGCAACCGTTATGTATTAATAAGCCAGTCTCTCC 1260
QY      421  GlyGluPheHisAlaGlnProAspAsnGlyThrPheGluLeuPheLeuLysGlyArgAsn 440
Db      1261  GGAGAAATTTTCATCCCGCGGATACAGGGACTTTTGAACTTTTATATAAGGGCAGAAAC 1320
QY      441  PheThrProAspAlaGlyValPheValTyrSerGlyAspGluAlaLeuMetLysLeuArg 460
Db      1321  TTTACCCAGACCCCGGGTATTGTGTATAGCGGCGACGAACCCATCATGAACACTGCGG 1380
QY      461  AsnTrpTyrArgGlnThrArgIleHisSerThrLeuThrLeuAspAsnGlnAsnMetVal 480
Db      1381  AACTGTGTACGTCACAAACCCGATACACAGCAGCGCTTACACTCGACAAATCAAAATATGGTC 1440
QY      481  IleThrLysAlaArgGlnAsnLysTrpGluThrGlyAsnAsnLeuAspValLeuThrTyr 500
Db      1441  ATTACCAAGCCCGGCAAAACAATGGGAACAGGAATAACCTTGATGTGCTTACCTAT 1500
QY      501  ThrAsnProSerTyrProAsnLeuAspHisGlnArgSerValLeuPheLeuAsnLysLys 520
Db      1501  ACCAACCCAGCTATCGAATCTGGACCATCAGCGCAGTGTACTTTTCATCAACAAAAA 1560
QY      521  TyrPheLeuValIleAspArgAlaIleGlyGluAlaThrGlyAsnLeuGlyValHisTrp 540
Db      1561  TACTTCTCGTCATCGATAGGCAATAGCGAAGCTACCGGAACCTGGCGGTACACTGG 1620
QY      541  GlnLeuLysGluAspSerAsnProValPheAspLysThrLysAsnArgValTyrThrThr 560
Db      1621  CAGCTTAAAGAGACAGCAACCTGTTTTGCGATAAGACAAGAACCGGGTTTACACCACT 1680
QY      561  TyrArgAspGlyAsnAsnLeuMetIleGlnSerLeuAsnAlaAspArgThrSerLeuAsn 580
Db      1681  TACAGAGATGGTAAACACCTGATGATCCATCTGTAATCGGACAGGACCGCCCTCAAT 1740
QY      581  GluGluGluGlyLysValSerTyrValTyrAsnLysGluLeuLysArgProAlaPheVal 600
Db      1741  GAAGAAGAGGAAAGGTACTTATGTTTACAAATAGAGAGCTGAAAGAGCTGCTTCGTA 1800
QY      601  PheGluLysProLysLysAsnAlaGlyThrGlnAsnPheValSerIleValTyrProTyr 620
Db      1801  TTTGAAAGCCCTAAAGAAAGTACCGGCACACAAAAATTTGTCAGTATAGTTTATCCATAC 1860
QY      621  AspGlyGluLysAlaProGluIleSerIleArgGluAsnLysGlyAsnAspPheGluLys 640
Db      1861  GACGGCCAGAGGCTCCAGAGTACGATACGGAATACCGGAACCAAGGCAATGATTTTGAGAA 1920
QY      641  GlyLysLeuAsnLeuThrLeuThrIleAsnGlyLysGlnGlnLeuValLeuValPro 659
Db      1921  GGCAAGCTTAACTTAACCTTACCATTAACGGAACACACAGCTTGTGTGTTCT 1977

RESULT 2
I71365
LOCUS      I71365
DEFINITION Sequence 3 from patent US 5681733.
ACCESSION I71365
VERSION    I71365.1 GI:3007500
KEYWORDS
SOURCE
ORGANISM   Unknown.
REFERENCE  1 (bases 1 to 1980)
AUTHORS   Su,H., Blain,F., Bennett,C., Gu,K., Zimmermann,J. and Musil,R.
TITLE      Nucleic acid sequences and expression systems for heparinase II and
            heparinase III derived from Flavobacterium heparinum
JOURNAL    Patent: US 5681733-A 3 28-OCT-1997;
FEATURES   Location/Qualifiers
            source
            1..1980
            /organism="unknown"
BASE COUNT 629 a 432 c 439 g 480 t
ORIGIN
Alignment Scores:

```

```

Pred. No.: 1 85e-280 Length: 1980
Score: 3494.00 Matches: 659
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

```

US-09-802-285A-2 (1-659) x I71365 (1-1980)

```

QY      1  MetThrThrLysIlePheLysArgIleIleValPheAlaValIleAlaLeuSerSerGly 20
Db      1  ATGACTACGAAATTTTAAAGGATCATTTGCTGTAATTCCTATCGTCGGGA 60
QY      21  AsnIleLeuAlaGlnSerSerIleThrArgLysAspPheAspHisIleAsnLeuGlu 40
Db      61  AATATATCTTGCACAAAGCTCTTCATTACCGAAGAGATTTTGACCACATCAACCTTGAG 120
QY      41  TyrSerGlyLeuGluLysValAsnLysAlaValAlaAlaGlyAsnTyrAspAlaAla 60
Db      121  TATTCGGACTGGAAAGGTTAATAAGCAGTTGCTCCGCGCAACTATGACGATGGGCC 180
QY      61  LysAlaLeuLeuAlaTyrTyrArgGluLysSerLysAlaArgGluProAspPheSerAsn 80
Db      181  AAAGCATTACTGGCATACTACAGGGAATAAGTAGGCCAGGAACTGATTCAGTAAT 240
QY      81  AlaGluLysProAlaAspIleArgGlnProIleAspLysValThrArgGluMetAlaAsp 100
Db      241  GCAGAAAGCCCTGCCGATATACGCCAGCCATAGATAGGTTACGCGTCAAAATGGCGAC 300
QY      101  LysAlaLeuValHisGlnPheGlnProHisLysGlyTyrGlyTyrPheAspTyrGlyLys 120
Db      301  AAGGCTTTGGTCCACAGTTTCAACCGCAAGGCTACGGCTATTTTATGTTGTAAT 360
QY      121  AspIleAsnTrpGlnMetTrpProValLysAspAsnGluValArgTrpGlnLeuHisArg 140
Db      361  GACATCAACTGGCAGATGTGGCGGTAAAGACAATGAAGTAGCTGCGTGCGAGTGCACCGT 420
QY      141  ValLysTrpTrpGlnAlaMetAlaLeuValTyrHisAlaThrGlyAspGluLysTyrAla 160
Db      421  GTAAATGTGTCAGCTATGCGCTGTTTATCAGCTACGCGCGCATGAAAAATATGCA 480
QY      161  ArgGluTrpValTyrGlnTyrSerAspTrpAlaArgLysAsnProLeuGlyLeuSerGln 180
Db      481  AGAGAAATGGGTATATCAGTACAGCATTCGGCCAGAAAAACCCATGCGCGCTGTCCGAG 540
QY      181  AspAsnAspLysPheValTrpArgProLeuGluValSerAspArgValGlnSerLeuPro 200
Db      541  GATAATGATAAATTTGTGTGCGCGCCCTTGAAGTGTGCGACAGGGTACAAGTCTTCCC 600
QY      201  ProThrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGluPhe 220
Db      601  CCAACCTTCAGCTTATTTGTAACACTCGCAGCGCTTTTACCCAGCGCTTTTAAATGGAATTT 660
QY      221  LeuAsnSerTyrHisGlnGlnAlaAspTyrLeuSerThrHisTyrAlaGluGlnGlyAsn 240
Db      661  TTAACAGTTTACCAACACAGCGCCGATTTATTCAGCATTTATCGCCAAACAGGGAAC 720
QY      241  HisArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPheLys 260
Db      721  CACCGCTTTATTGAAGCCCAACGCAACTGTTTGCAGGGGTATCTTCCCTGAAATTAAT 780
QY      261  AspSerProArgTrpArgGlnThrGlyIleSerValLeuAsnThrGluIleLysGln 280
Db      781  GATTCCACCAAGATGAGGCAACCGGCATATCGGTGCTGAACACCGCAGATCAAAAACAG 840
QY      281  ValTyrAlaAspGlyMetGlnPheGlnLeuSerProIleTyrHisValAlaAlaIleAsp 300
Db      841  GTTTATGCCGATGGGATGCAGTTTGAACCTTTCCACAAATTTACCATGTAGTCCCATCGAT 900
QY      301  IlePheLeuLysAlaTyrGlySerAlaLysArgValAsnLeuGluLysGluPheProGln 320
Db      901  ATCTTCTTAAGGCCCTATGGTTCTCCAAACAGATTAACTTGAAGAAAGAAATTTCCGCA 960

```

```
Qy 321 SerTyrValGlnThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuProAsp 340
Db 961 TCTATGACAACTGTGAAAAATATGATTATATGGCGCTGATCAGTATTTTCACTGCCAGAT 1020
Qy 341 TyrAsnThrProMetPheGlyAspSerTrpIleThrAspLysAsnPheArgMetAlaGln 360
Db 1021 TATAACACCCCTATGTTGGAGATTTCATGGATTACAGATATAAAATTTTCAGGATGCCACAG 1080
Qy 361 PheAlaSerTrpAlaArgValPheProAlaAsnGlnAlaIleLysTyrPheAlaThrAsp 380
Db 1081 TTTGCCAGCTGGCCCGGTTTCCCGCAACACCGGCATATAAAATATTTTGTCTACAGAT 1140
Qy 381 GlyLysGlnGlyLysAlaProAsnPheLeuSerLysAlaLeuSerAsnAlaGlyPheTyr 400
Db 1141 GGCACACAGGTAAAGCGCTTAACCTTTTATCCAAAGCATTCAGCAATGCAGGCTTTTAT 1200
Qy 401 ThrPheArgSerGlyTyrAspLysAsnAlaThrValMetValLeuLysAlaSerProPro 420
Db 1201 ACGTTTAGAAGCGGATGGGATAAAAAATGCAACCGTTATGGTATTAAAAAGCCAGTCTCC 1260
Qy 421 GlyGluPheHisAlaGlnProAspAsnGlyThrPheGluLeuPheIleLysGlyArgAsn 440
Db 1261 GGGGAATTTTCAGCCCGCAGCGGATACCGGACTTTTGAACCTTTTATAAGGGCGAGAAC 1320
Qy 441 PheThrProAspAlaGlyValPheValTyrSerGlyAspGluAlaIleMetLysLeuArg 460
Db 1321 TTTACCCAGAGCGCGGGTATTTGTGTATAGCGCGAGCCATCATGAACTGGG 1380
Qy 461 AsnTrpTyrArgGlnThrArgIleHisSerThrLeuThrLeuAspAsnGlnAsnMetVal 480
Db 1381 AACTGGTACCGTCAAAACCGCATACACAGCAGCTTACATCGACCAATCAAAATATGGTC 1440
Qy 481 IleThrLysAlaArgGlnAsnLysTrpGluThrGlyAsnAsnLeuAspValLeuThrTyr 500
Db 1441 ATTCAAGACCGCGCAAAACAAATGGGAAACAGGAAATAACCTTGATGTGCTTACCAT 1500
Qy 501 ThrAsnProSerTyrProAsnLeuAspHisGlnArgSerValLeuPheIleAsnLysLys 520
Db 1501 ACCAACCCAGCATTCGGAATCTCGACCATCAGCGCAGTGACTTTTTCATCAACAAAAA 1560
Qy 521 TyrPheLeuValIleAspArgAlaIleGlyGluAlaThrGlyAsnLeuGlyValHisTrp 540
Db 1561 TACTTTCTGGTCATCGATAGGCAATAGCGGAAGCTACCGGAACCTGGCGGTACACTGG 1620
Qy 541 GlnLeuLysGluAspSerAsnProValPheAspLysThrLysAsnArgValTyrThrThr 560
Db 1621 CAGCTTAAAGAAAGCAGCAACCTGTTTTCGATAAGACAAAGACCGGGTTTACACCAT 1680
Qy 561 TyrArgAspGlyAsnAsnLeuMetIleGlnSerLeuAsnAlaAspArgThrSerLeuAsn 580
Db 1681 TACAGAGATGGTAAACACCTGATGATCCAATCGTTGAATGGGACAGGACAGCCTCAAT 1740
Qy 581 GluGluGluGlyLysValSerTyrValTyrAsnLysGluLeuLysArgProAlaPheVal 600
Db 1741 GAAGAAGAAGGAAAGAGTATCTTATGTTTACATAAGAGAGCTGAAAGACCTGCTTCGTA 1800
Qy 601 PheGluLysProLysLysAsnAlaGlyThrGlnAsnPheValSerIleValTyrProTyr 620
Db 1801 TTTGAAAGACCTPAAAAAGATGCCCGCACACAAAATTTTGTGATATAGTTATTCATAC 1860
Qy 621 AspGlyGlnLysAlaProGluLysSerIleArgGluAsnLysGlyAsnAspPheGluLys 640
Db 1861 GACGCCCAAGAGCTCCAGAGATCAGCATACGATACGGAAGAAACAAAGGCAATGATTTTCAGAAA 1920
Qy 641 GlyLysLeuAsnLeuThrLeuThrIleAsnGlyLysGlnGlnLeuValLeuPro 659
Db 1921 GGCAAGCTTAATCTAACCCCTTACATTAACGGAAACACACAGCTTGTGTGTTGTTCT 1977
RESULT 3
AE014145/c
LOCUS
DEFINITION Streptococcus pyogenes MGAS315, section 10 of 37 of the complete genome.
```

```
ACCESSION AE014145 AE014074
VERSION AE014145.1 GI:21904143
SOURCE
ORGANISM Streptococcus pyogenes MGAS315.
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.
REFERENCE 1 (bases 1 to 50327)
AUTHORS Beres,S.B., Sylva,G.L., Barbican,K.D., Lei,B., Hoff,J.S., Mammarella,N.D., Liu,M.-Y., Smoot,J.C., Porcella,S.F., Parkins,L.D., McCormick,J.K., Leung,D.Y.M., Schlievert,P.M. and Musser,J.M.
TITLE Genome sequence of a serotype M3 strain of group A Streptococcus: Phage-encoded toxins, the high-virulence phenotype, and clone emergence
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (15), 10078-10083 (2002)
REFERENCE 2 (bases 1 to 50327)
AUTHORS Beres,S.B., Sylva,G.L., Barbican,K.D., Lei,B., Hoff,J.S., Mammarella,N.D., Liu,M.-Y., Smoot,J.C., Porcella,S.F., Parkins,L.D., McCormick,J.K., Leung,D.Y.M., Schlievert,P.M. and Musser,J.M.
TITLE Direct Submission
JOURNAL Submitted (14-JUN-2002) Laboratory of Human-Bacterial Pathogenesis, Rocky Mountain Laboratories, NIAID, NIH, 903 South Fourth St., Hamilton, MT 59840, USA
FEATURES
source
1..50327
/organism="Streptococcus pyogenes MGAS315"
/strain="MGAS315"
/serotype="M3"
/db_xref="taxon:198466"
/notes="group: A"
95..307
/gene="SpyM3_0419"
95..307
/gene="SpyM3_0419"
/notes="best non-GAS blastp hit: gb|AAK74919.1| (AE007385) conserved hypothetical protein (Streptococcus pneumoniae TIG84), and gb|AAK99494.1| (AE008446) Conserved hypothetical protein (Streptococcus pneumoniae R6)"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAW79026.1"
/db_xref="GI:21904144"
/translation="MAEFTFNEEHLTLSENDKGWTKELNRSVNGAEAKWDIRTS PDHFMGIGITLNEEFTLIDAFKK"
complement (686..2179)
/gene="lyss"
/notes="SpyM3_0420"
complement (686..2179)
/gene="lyss"
/notes="best blastp hit: gb|AAK33574.1| (AE006515) putative lysyl-tRNA synthetase (Streptococcus pyogenes M1 GAS)"
/codon_start=1
/transl_table=11
/product="putative lysyl-tRNA synthetase"
/protein_id="AAW79027.1"
/db_xref="GI:21904145"
/translation="MSNQHIEELNDQOIVREKMMALAEQQIDPPGKFRGTANSDEL KENADTKKEELHELNETAIVAGRLMTXRGKGVGFALQDREGQIQLYVRKDSYGED NYEIPKADLGDFTGVEGVMRTDMGELSIKATKLHLSKSLPLPERFHLGDTETI YRKRLDLISNRESGDFRVTSRKMSIERRYLDGLDPLEVTPVLHNAGCAAAAPFV THNAQNDLMLRLATLHLKELVGGMERYEIGRIERNESGMDATHNPEITSIVYQ AYAYLDIMNLTEGLIQHAAKAVKGDGPDYQSTEIRINEPFRKRVHMDA.KENTGVQ FWPENVEEALAKEQVPLEKFTSVGHILINAFEEFVEELVQPTFVFGHPHEVS PLAKNPEDTEFTDFELFMKYEYANAFELNDPFDQLSRFAQAQAKELGDDDEATG IDYDFVEALREYCMPTTGLGIDRLCMLLTNTTTIRDLVLLFPTMKP"
2353..3255
/gene="SpyM3_0421"
2353..3255
/gene="SpyM3_0421"
/notes="best non-GAS blastp hit: emb|CAC98713.1| (AJ591976)
```

```

lmo0635 [Listeria monocytogenes]
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/db_xref="GI:21904146"
/transl_table=11
/db_xref="GI:21904147"
/translation="MKLYFVRHCKTLWNLEGRFGAGGSPILLEAKDEIHLGKELA
KVAFDAVYSDIORAMATAAILDADFQPKLXHTQDLREWRGLKLGAKIATMAIY
PQOMLAFRENLAQFKDFEASESIYQTORVCHLIQSFKDXYONVLIIVGHGANLTAT
IRSLGFEPALLAKGLDNTSLTILETKDYLYDCLINWDRKSFIEV"
/complement(4327..4806)
/gene="SpyM3_0423"
/complement(4327..4806)
/gene="SpyM3_0423"
/translation="best non-GAS blastp hit: gb|AAK75104.1| (AE007402)
regulatory protein, putative [Streptococcus pneumoniae
TIGR4]"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="AA079030.1"
/db_xref="GI:21904148"
/translation="MAKTKLKKTLVEQLDKRANIAHQGLKLNALRGDFPDQLPSDI
XYTLALTGQTPGLIGLIPLETHLSEKOLAKSGNKKVSMVPKDLQKTYTHGANN
PVGIROKHSYPFIDOTALKGGQIIIVSAGEVGRSIISSQALADFVGSFADLKKRK"
/complement(4896..5459)
/gene="SpyM3_0424"
/complement(4896..5459)
/gene="SpyM3_0424"
/translation="best non-GAS blastp hit: emb|CAC99507.1| (AL591979)
similar to unknown proteins [Listeria monocytogenes]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA079031.1"
/db_xref="GI:21904149"
/translation="WSPNTNVKYLIEAALFAALMALSFIPDPAGWFSPYGAIALVI
FSLRLKGLMAGLWGLHVLGVKGYLSQVFEIYLAFTSGLGASFSLSLK
TLRRQOTPFVAFVIAWASLAVTVRLVHFLAGLIFWGSYAPGSAVWYSPVNGTA
GVLTFLITCLAMIALPIHPQLFPKD"
/complement(5728..6576)
/gene="SpyM3_0425"
/complement(5728..6576)
/gene="SpyM3_0425"
/translation="best blastp hit: gb|AAK33579.1| (AE006515) putative
endolysin, phase associated [Streptococcus pyogenes M1
GAS]"
/codon_start=1
/transl_table=11
/product="putative endolysin, phase associated"
/protein_id="AA079032.1"
/db_xref="GI:21904150"
/translation="WRRKIKPIVILVFFILLAMVLIIGKROANHAQKEVEDAKSHIP
IATSNPSKARTSTETEDFILNPVDVSGQLPEEIDYDLSRHISGAIVRYGSGQI

```

```

TAHNAFTTIDKSPKTHIKERQKNVPVAVYSYALGRSTKEMKEBARAFYKNAAPY
NPYYIWDVEETAKMDKNMGVTAFREELKGLGAENGLYIGTFMAEQDILSTKGFDAV
WIPTGSDSGSYEAAPNTTLDYDLHQYTSQYLSGFNNALDLNQIAVTOKTKTFFEL
FGNSNN"
/gene="SpyM3_0426"
/translation="MSAKFMSRIAMSALCVLRMVFIISLPNVQVPTAFLLSYLVFG
LAEAVLVMCLFLSAFLIGRGPWFQVTCFVLVLLWRFVLYPLSQQFPKYOLGCO
AELVALCGLLYGVLIDTCFAYLYSMFWSYVLGMPFNIAHALSTLVFVPMVLMFRF
LISEQKI"
/translation="best non-GAS blastp hit: gb|AAK05269.1| (AE006349_4
(AE006349) unknown protein [Lactococcus lactis subsp.
lactis]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA079033.1"
/db_xref="GI:21904151"
/translation="MNRKFTGTLALVAMLVACSQGTQOTPSVPKADHHVRLVVK
RBTNIDKVSFGKGTVEILKDNVEYKEDGFTIADIGEODTKANKYWLFPKNGK
MADKADQITVKDGSISFYQVEFK"
/complement(7817..8296)
/gene="SpyM3_0428"
/complement(7817..8296)
/gene="SpyM3_0428"
/translation="best blastp hit: gb|AAK33582.1| (AE006515) putative
glutathione peroxidase [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative glutathione peroxidase"
/protein_id="AA079035.1"
/db_xref="GI:21904153"
/translation="MNPVLYDFTVKAQDSDISLVNYKVKVWLVTATKGLTPQYOA
LQALYDTHDKGFVLDPPCNQTPCEAEIINRFCSLYHTTTPERAKLVNGKD
ADPLFTWLKEKSGPLGKRIEWNFTKFLIDQGVQIKRYSKTDPKLIEDLKALLG"
/complement(8289..10088)
/gene="pepf"

```

Alignment Scores:

Pred. No.:	Length:	Matches:
Score:	2,63e-08	50327
Percent Similarity:	235.50	152
Best local Similarity:	36.27%	Conservative: 87
Query Match:	23.07%	Mismatches: 282
	6.74%	Indels: 140
	1	Gaps: 26

US-09-802-285A-2 (1-659) x AE014145 (1-50327)

QY 64 LeuAlaTyrrArgGluLysSerLysAlaArgGluProAspPheSer-----79

Db 28575 CTTGCTTTCGACGCTTTAAAGAAACAGTC---AATCCAGACTTTTGCAGAAATACCTG 28519

QY 80 -----AsnAlaGluLysProAlaSplleArgGlnProIleAspLysValThrArg 96

Db 28518 CTAGACTATCAGGCTAACAGCTATGCTGAT-----CAAAAA 28483

QY 97 GluMetAlaAspLysAlaLeuValHisGln-----PheGlnProHisLysGlyTyrGly 114

Db 28482 AGAATAGCTGACCTACTCTTAACCAATACCTTTTATTGAGATAATTG-GGATATGGA 28424

```

115 TyrPheAspTyrGly-----LysAsp-IleAsnTrpGlnMetTrpProValIysAs 131
Db 28423 GCCCTGCTCTATTCCTATCGTTGGAGACTATCACTTGGCAAAAGCGCGTGACAGATGA 28364
Qy 131 pAsnGluValArgTrpGlnLeuHisArgValIysTrpTrpGlnAlaMetAlaLeuValIys 151
Db 28363 TCCTGAGTGAACCTTATCTTAATCGACAGACCTACCTCCAAACATTAACTCTCGTCTA 28304
Qy 151 rHisAlaThrGlyAspGluLysTyr-----AlaArgGluTrpValTyrGlnTyrSe 168
Db 28303 CCTTGTGAAGACAGATGAACGGTATCTCTTAACAGCTAAAGGTTCATCTCAATTGGAT 28244
Qy 168 rAspTrpAla---ArgLysAsnProLeuGlyLeuSerGlnAspAsnLysPheValTr 187
Db 28243 TGAATCAGCTATTCCTTTAGACCTTAAGGGTTAGCTACTAG-AACCTTAGATAC---TG 28188
Qy 187 pArgProLeuGluValSerAspArgValGlnSerLeuPro-ProThrPheSerLeuPheV 207
Db 28187 GCATCCGTTGTTTGTCTGGGTCAAATGTTTGAATTAACCTTAATCTTTTAAAC----- 28135
Qy 207 alAsnSerProAlaPheThrProAlaPheLeuMetGluPheLeuAsnSerTyrHisGlnG 227
Db 28134 -----GCCCTTAACCAACAAAGAGGAAGCCTCATCTAGCTTCCCTCGGAAAAAC 28086
Qy 227 InAlaAspTyrLeuSerThrHisTyrAlaGluGln-----GlyAsnHisArgLeuP 244
Db 28085 AACTTCATTTCTCATACCACTACTCTGTGATATACAGCTCAGCAATCGGGATCT 28026
Qy 244 heGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPheLysAspSerProA 264
Db 28025 TGCAAACCACTCGCATTTTATTAGCTGATGCTACTACTGTTCTGACCTTGATATTGCTG 27966
Qy 264 rGTrpArgGlnThrGlyIleSerValLeuAsnThrGluIleLysIysGlnValTyrAlaA 284
Db 27965 CAGCAACACCTTTCGCTAGAAAAGAAATTGACCAACAAATCGCACTACAAATTCCTGAAG 27906
Qy 284 spGlyMetGlnPheGluLeuSerProIleTyrHisValAlaIleAspIlePheLeuL 304
Db 27905 ACGGTAGTCAATTTGACACATCAACCATGATGATGAGGTGTAAATCCCTTTTGG 27846
Qy 304 yAlaTyrGlySerAlaLysArgValAsnLeuGluLysGluPheProGlnSerTyrValG 324
Db 27845 AATTA-----ACTGCTCTAGTCCCTGATTATCTTCCACAGCTTAGAC 27804
Qy 324 InThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuProAspTyrAsnThrP 344
Db 27803 CTACTTTGCTTCCCATGAGTGAATATCTCCTCAAAATGATGATCCTGACCAACAAACAGA 27744
Qy 344 roMetPheGlyAspSerTrpIleThrAspLysAsnPheArgMetAlaGlnPheAlaSer- 363
Db 27743 TCCCTTGGAGATAGTATGATTTTACGATACCAAGATATTTAACTAGCAGCCACTA 27684
Qy 364 -----TrpAlaArgValPheProAlaAsnGlnAlaIleLysTyrP 377
Db 27683 TCCTTTGAGAACCAACACACTTAAAGGTGCTGCTTTTCCACACTTGAT---ATCGATAGTC 27627
Qy 377 heAlaThrAspGlyLysGlnGly-----LysAlap 387
Db 27626 TTCTCCTTCTAGGAAAGAGGTGTCACACTTTTGAACAGTTTACCCATCCAAACCTTAC 27567
Qy 387 roAsnPheLeuSerLysAlaLeuSerAsnAlaGlyPheTyrThrPheArgSerGlyTrpA 407
Db 27566 CAACTTTT---GGCATCATCTTTGAACATTCGGGGACATTAACATC-----A 27522
Qy 407 spLysAsnAlaThrValMetValLeuLysAlaSerProProGlyGluPheHisAlaGlnP 427
Db 27521 ATCAAGAAATATTATCTCTTTTTCAAAATATGGGCCATCGCTAGTTCCCATACCCATA 27462
Qy 427 roAsnAsnGlyThrPheGluLeuPheIleLysGlyValArgAsnPheThrProAspAlaGlyV 447
Db 27461 GCGACCAAAATAGTATCTGCTCTATATATAGGACCAACCACTTTTCTGTGATGCTGGGC 27402
447 alpPheValTyrSerGlyAspGluAlaIleMetLysLeuArgAsnTrpTyrArgGlnThrA 467
Db 27401 GATACACTTACAAAGAGAA-----CCATACGCTATGCCCTAANAAGCGCCA 27354
Qy 467 rGileHisSerThrLeuThrLeuAspAsnGlnAsnMetValIleThrLysAlaArgGlnA 487
Db 27353 GCCATCAATCTACCGCTTCTTAGAAGAACAA-----TTACCTGAGCAGATTGATA 27303
Qy 487 snLysTrpGluThrGlyAsnAsnLeuAspValLeuThrTyrThrAsnProSerTyr- 505
Db 27302 GCTCTCTGGGCT-----TACCTTTCTTACCCCAATCCAAATTATGTC 27261
Qy 505 ----- 505
Db 27260 ATCTTCGACAAAATGACACAGCTTTACTTTATTGATGGTAGCTACACAGCGCAATTCCTG 27201
Qy 506 -----ProAsnLeuAspHisGlnArgSerValLeuPheIleAsnLysIysTyrPheLeuV 524
Db 27200 ATAGAAACAACACTTACCAACATGACCGCAAAATCTCATCTTACCACAGCAATTTTCTTA 27141
Qy 524 alIleAspArgAlaIleGlyGluAlaThrGlyAsn-----L 536
Db 27140 TCATTTGATACGATA-----CAGGCTCAAGGCAACCACTGTTGTCAGTCAATTTATTC 27087
Qy 536 euGlyValHisTrpGlnLeuLysGluAspSerAsnProValPheAspLysThrLysAsnA 556
Db 27086 TTGATAANTCACTTAGAAGTAAAGACAGAT-----CACCTTAGTGACCTCA 27042
Qy 556 rGValTyrThrTyrArgAspGlyAsnAsnLeuMetIleGlnSerLeuAsnAlaAspA 576
Db 27041 GACTAATCAGTGACTAC----- 27025
Qy 576 rGThrSerLeuAsnGluGluGlyLysValSerTyrValTyrAsnLysGluLeuLysA 596
Db 27024 -----CCTTCTACTATAGAAAGACGATTTTGTCAAAAATAACACCACTACCTGACTA 26970
Qy 596 rGProAlaPheValPheGluLysProLysLysAsnAlaGlyThrGlnAsnPheValSerI 616
Db 26969 GTCACAACTCATCAACAAACGAAACCTTTTAAAGACAGAGGTTCACCACT-----ACCC 26916
Qy 616 leValTyrProTyrAspGlyGlnLysAlaProGluIleSerIleArgGluAsnLysGlyA 636
Db 26915 TGTTAGTCCCTGACGATACCAAGTTACTCCCTTAACACCACTTCAACAGCAACGCA 26856
Qy 636 snAspPheGluLysGlyLysLeuAsnLeuThrLeuThrIleAsnGlyLysGln 653
Db 26855 ACCCAATTGAGACCGCC-----CTTCTCTGCATTTAAAGAAAGCAG 26812
RESULT 4
AE006517/c 12195 bp DNA linear BCT 01-JUN-2001
LOCUS Streptococcus pyogenes M1 GAS strain SF370, section 46 of 167 of
DEFINITION the complete genome.
ACCESSION AE006517 AE004092
VERSION AE006517.1 GI:13621815
KEYWORDS
SOURCE Streptococcus pyogenes M1 GAS.
ORGANISM Streptococcus pyogenes M1 GAS.
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
1 (bases 1 to 12195)
Ferrretti, J.J., McShan, W.M., Adjic, D., Savic, D., Savic, G., Lyon, K.,
Primeaux, C., Sezate, S.S., Surorov, A.N., Kenton, S., Lai, H., Lin, S.,
Qian, Y., Jia, H.G., Najjar, F.Z., Ren, Q., Zhu, H., Song, L., White, J.,
Yuan, X., Clifton, S.W., Roe, H.A. and McLaughlin, R.E.
Complete genome sequence of an M1 strain of Streptococcus pyogenes
Proc. Natl. Acad. Sci. U.S.A. 98 (8), 4658-4663 (2001)
2 (bases 1 to 12195)
Ferrretti, J.J., McShan, W.M., Adjic, D., Savic, D., Savic, G., Lyon, K.,
Primeaux, C., Sezate, S.S., Surorov, A.N., Kenton, S., Lai, H., Lin, S.,
Qian, Y., Jia, H.G., Najjar, F.Z., Ren, Q., Zhu, H., Song, L., White, J.,

```

Yuan, X., Clifton, S.W., Roe, B.A. and McLaughlin, R.E.

TITLE JOURNAL

Direct Submission
Submitted (10-APR-2001) Department of Microbiology and Immunology,
University of Oklahoma Health Sciences Center, 940 SL Young Blvd.,
Oklahoma City, OK 73104, USA

FEATURES

```

source
1. .12195
/organism="Streptococcus pyogenes M1 GAS"
/strain="SF370"
/serotype="M1"
/db_xref="taxon:160490"
complement (106. .339)
/gene="SPY0619"
complement (106. .339)
/gene="SPY0619"
/notes="Best Blastp hit = no blast hits"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK33591.1"
/db_xref="GI:13621816"
/translations="MEFYALFLSLGSKFRVTSPLVLTPLLPOQKAVALAHVTRNNK
SVKSAEQMLFPHQTFGVSLRLKKEYFLQSP"
complement (411. .1712)
/gene="SPY0621"
complement (411. .1712)
/gene="SPY0621"
/notes="Best Blastp hit = sp|P39651|YWFO_BACSU HYPOTHETICAL
51.0 KD PROTEIN IN PTA 3'REGION >gi|7450837|pir||G70056
YWFO protein - Bacillus subtilis
>gi|1561567|emb|CAB02494.1| (Z80355) Unknown [Bacillus
subtilis] >gi|2636296|emb|CAB15787.1| (Z99123) alternate
gene name: ipa-93d
similar to hypothetical proteins [Bacillus subtilis]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK33592.1"
/db_xref="GI:13621817"
/translations="MNEKFRDVRVNYIHDNPLIYDLINTKEFQRLRIKQVPTTAF
TFGADHSRFLSGVEITARRVTAIFEEKYADIMNKDSLTMTAALHDIHGHAYS
HYFEVLTHDFAETQEIITNPETAINILVRHAPDFPKVASVINHTYPNKOVVOLI
SSQICDMDYLLRDSYFASAGQFDLRLVIRPBDGIVFHSGHMVAHEDYIVS
RFQVMQDVYFHPASRAVELLQNLKRAOHLVPEQAYFQKTAPGLIPFEKKANLAD
YIALDDGVNTPQVMASEDHLSLASEFINRKLKSVTFDQDSQGBLERLRQIWE
SYGDPDXYTGTHINFDLPDIYRPLENFRQIENMQDGSIAELSQLSPIVKAULTG
TYGDRRRFFPKMELEDDLFLAPSKETFMYSYISNGHFHSQ"
1794. .2180
/gene="SPY0622"
1794. .2180
/gene="SPY0622"
/notes="Best Blastp hit = no blast hits"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK33593.1"
/db_xref="GI:13621818"
/translations="MKLQTNIRFGDETEIIQEIHDCEWREKGGVQYLIYQNTDKEK
VIKNETELTSRFSNPQSIMKFFAGKRVLIATLPTMGVQOQFLDITSHYLDSCQK
LDLHYLLQAOQEMLFASYHLELSWD"
2411. .5092
/gene="pacl"
/notes="SPY0623"
2411. .5092
/gene="pacl"
/EC_number="3.6.1.-"
/notes="Best Blastp hit = pir||H69877 calcium-transporting
ATPase homolog ylob - Bacillus subtilis
>gi|2337795|emb|CAA74269.1| (Y19337) putative Pacl protein
[Bacillus subtilis] >gi|2633938|emb|CAB13439.1| (Z99112)
similar to calcium-transporting ATPase [Bacillus
subtilis]"
/codon_start=1

```

```

/transl_table=11
/product="putative calcium-transporting ATPase"
/protein_id="AAK33594.1"
/db_xref="GI:13621819"
/translations="MSKEQRHEAFYTOSEETVLAQLETSEGLTSAQAKERLAEGYGRN
ELDEGRSRSLFMKLFQFDLMIILIIAALISVLTEGEGLTDAIILAVILNRAAP
GYQEQAARAAIEALKSMSPLEARIRRDGHVTEIDSKELVPGDIVILLEAGDVPADUR
LLEANSKLEEARALTGESVFEKDLSTAVESDAGIDRVNMGYQNSNVTVGRGIVGIT
NTEMYTEVGHIAGMANADETTPLEKNDLNSKILTYAILVIAAVTFAVFLRGQH
PLEGMTSVALAAVAIEPEGLPAIVTVLSLGTQVLAKRNAIRKLPATEVLTGSTETIA
SDKTGLTMNOMTEKVTNGTNGLOSSADIAFDNTTLRVNMFANDTKVDSGKLIIGDP
TETALVEGLDRNFVREMAVEPRVAELPEDSRDKLMSTIHKAQDGFIAPVKGAPD
QALLKRVTOLENGQIRPITDADKTKTILDTNSLAKALRULMAYKYSALPTLETETI
VEANVTSGLVGMIDPERPEAAQAVAKAEGIRPIMTGDHQTDAKAKRGLIIEE
DGDVHVTGAELNELSEDFQVFKQYSVYARVSPHEKVIKVAWEGEKVAMTGDG
VNDPSLKTADIGMIGITGVTEVSKGASMDVADNDFATIIIVAVEGRKVFSNIQTI
PRGRSSFFDGGVKRAILYQGAFTILVGVGFALMFPEHTSYHDVHADALTMAYVT
LGLIQLAHAYNVKSUYQSIPTVGLPKNKLPNYSIPVAFVALMATVTVFGNQFFHVT
LITQMLVIIIGSLMVLVBLVKAQVRSLSQDEKAI"
complement (5176. .6171)
/gene="regr"
/notes="regr"
complement (5176. .6171)
/gene="regr"
/notes="Best Blastp hit = sp|P50844|KDGR_BACSU KDG OPERON
REPRESSOR >gi|7442871|pir||B69648 transcription repressor
of pectin utilization operon kdgr - Bacillus subtilis
>gi|1146188|gb|AAB38478.1| (U47838) repressor protein
[Bacillus subtilis] >gi|2634631|emb|CAB14129.1| (Z99115)
transcriptional regulator (Laci family) [Bacillus
subtilis]"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator (Laci
family)"
/protein_id="AAK33595.1"
/db_xref="GI:13621820"
/translations="MQRKVTIKDIAELAKTSKTVTFYVINGRFDKMSSEETKNRISSEI
KATNYPSPARSLNAKSKTLGVIGDITNSFNSQIVKGIESKAEQFYQIIGNSN
KATNYPDEDELEKMLGVDFGIQTPNSPKYSRIIDIKKKVVPFDSLVYERTHWVI
KNNYDADVDTTQCCIDKGVEHFIMTGNPNLLSTFEPASGFIDVLEANHLTHQMI
IDENOTSBAIAQFLQGSITKSLVFPNCWALPKVFTAMKSLKFNIPGLVGFENI
EWKFSPLTTITIPAVEGEQATKILIDDIIEGHSQEAQKQIFDCQVNWQESTP"
complement (6235. .8142)
/gene="SPY0628"
complement (6235. .8142)
/gene="SPY0628"
/notes="Best Blastp hit = pir||T36084 hypothetical protein
SCE134.02c - Streptomyces coelicolor
>gi|4678627|emb|CAB41200.1| (AL049661) hypothetical
protein [Streptomyces coelicolor A3(2)]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK33596.1"
/db_xref="GI:13621821"
/translations="MFFAPAREKTVNPDFCRNYLLDYQTSYADQKRIADLLLTNTF
LPEDNDMEPCHIPYHLDPIITWQEAVIDDPEWFMNLRQTYLQKLIIVLVVERDERYL
LTAKGFIINWIESAIPDKGLATRLDTGIRCFAWVKCLIVLNLNALTQKOEESLIL
ASLEKLOPLHANVLDKYSLNWGILOTTAILADAYPGSDLDIAATAPAKELTQQ
IALQILEGDSQFEQSTWVHEVLKALLELTALVPDYLQPLRPTLLAMSDYLKMTGPD
HKQIFIGSDSDVTDRDITLTAATILEEHLKAAAFPTDILDSLLGKRGVHTFQPLP
VQTLPTFAHFEHSGHITINQENYIFFKNGIPGSHSDQNSCLYKQGFPCDA
GRYTYKEEPLRYALKASGHSTAFLEQPEQIDSWAYLSYKPNCHLRONGHYVF
TEGSYQTOFSDRNNYCHDRQILILPGGFIILDTTQACQNHCLVSOQFILDNDLVKTD
HLSDLRLSDGCPTEETILSKKNQYLTSKHLIKRPFKDKGCTSTLLVPDQTKVP
LTPLOTGRNPETALSMLKQFDYSICVLQEDLIKGEKLVILNLSHRIKRVVIVIN
HITNFIILKXH"
complement (8229. .9050)
/gene="agaD"
/notes="SPY0629"
complement (8229. .9050)

```

```

/gene="agad"
/note="Best Blastp hit = sp|P42911|PTPD ECOLI PTS SYSTEM,
N-ACETYL GALACTOSAMINE-SPECIFIC IID COMPONENT (EIID-AGA)
(N-ACETYL GALACTOSAMINE-PERMEASE IID COMPONENT)
(PHOSPHOTRANSFERASE ENZYME II, D COMPONENT)
>gi|7434458|pit|H65103|pts system,
n-acetyl galactosamine-specific IID component - Escherichia
coli (strain K-12) >gi|1789529|gb|AAC76174.1| (AE000395)
PTS system, N-acetylglucosamine enzyme IID component 1
[Escherichia coli] >gi|18895756|gb|AAF81092.1|AF228498_12
[Escherichia coli] Agad [Escherichia coli]
/codon_start=1
/transl_table=11
/product="putative PTS dependent galactosamine IID
component"
/protein_id="AAK33597.1"
/db_xref="GI:13621822"
/translation="MTGSKLTKDYLTALRAFFLQNGFNNTYQGIGYANVIYPAL
KKHFGNDKGLYQALEDNCFYNTNPHLPFTLSHLVLENNRPEETRNKMLALMG
PLAGIGDSLSQFLAPLFTSTAAASLASGLVLGPILFLAMNIIITAIKIGSLGYK
VTSFDKLSQMAVVSMAIVGVTVLAGLAATSVKLTVPITPAAGKVDAAATQAF
VTIQMLDKIAPALPALFTLLMYLILKNKWTYKLVILTVIIGVIGSLGILA"
complement(9037..9819)
/gene="agaw"
/note="SPY0630"
/complement(9037..9819)
/gene="agaw"
/note="Best Blastp hit = gb|AAF81084.1|AF228498_4
(AF228498) Agaw [Escherichia coli]"
/codon_start=1
/transl_table=11

```

Alignment Scores:

```

Pred. No.: 4, 83e-09 Length: 12195
Score: 234.50 Matches: 140
Percent Similarity: 36.35% Conservative: 85
Best Local Similarity: 22.62% Mismatches: 273
Query Match: 6.71% Indels: 122
DB: 1 Gaps: 23

```

US-09-802-285A-2 (1-659) x AE006517 (1-12195)

```

QY 96 ArgGluMetAlaAspLysAlaLeuValHisGlnPheGlnProHisLysGlyTyrGly--- 114
Db 8046 AAAAGAAAGTAGTGACCTACTCTTAACCAATACCTTTTATTGAAGATAATGGGATAG 7987
QY 115 -----TyrPheAspTyrGlyLysAsp-----IleAsnTrpGlnMetTrpProValLys 130
Db 7986 GAGCCCTGCCACATCCCTTACCCTTAGATCCCATTTACTTGGCAAGAGCTGTCTATGAT 7927
QY 131 AspAsnGluValArgTrpGlnLeuHisArgValLysTrpTrpGlnAlaMetAlaLeuVal 150
Db 7926 GATCTGAGTGGAACTTATGTTAATCGACAGACCTATCTCCAAAATAATATCTCTGHC 7867
QY 151 TyrHisAlaThrGlyAspGluLysTyr-----AlaArgGluTrpValTyrGlnTyr 167
Db 7866 TACCTTGTGAAAGAGATGAACGGTATCTTAAACAGCTAAAGGTTTTCATCTCAATTGG 7807
QY 168 SerAspTrpAla---ArgLysAsnProLeuGlyLeuSerGlnAspAsnLysPheVal 186
Db 7806 ATGAATCAGCTATCTCTTTAGACCTTAAGGGTTAGCTACTAG-AACCTTAGATAC--- 7751
QY 187 TrpArgProLeuGluValSerArgValGlnSerLeuPro-ProThrPheSerLeuPh 206
Db 7750 TGGCATCCGTTGTTTGGTCAAAATGTTGATTACCTTAACTTTTAAAC----- 7696
QY 206 eValAsnSerProAlaPheThrProAlaPheLeuMetGluPheLeuAsnSerTyrHisG 226
Db 7695 -----GCCTTAACCAAGCAAGAGAAAGCCTCATCTAGCTTCCCTGGAAAA 7649
QY 226 nGlnAlaAspTyrLeuSerThrHisTyrAlaGluGln-----GlyAsnHisArgLe 243
Db 7648 ACAACTTCAATTTCTCCATGCCCACTACTCTTGATAAATACAGCTCAGCAATTTGGGAAT 7589

```

```

QY 243 upheGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPheLysAspSerPr 263
Db 7588 CTTGCAAAACCCTGCGATTTTATTAGCTGATGCGCTACTTTGGTTCTGACCTTGATATGC 7529
QY 263 oArgTrpArgGlnThrGlyIleSerValLeuAsnThrGluLysLysGlnValTyrAl 283
Db 7528 TGCAGCAACAGCTTTTCGTAGAAAAGAAATTGACGCAACAAATCGCACATACAAATCTCGA 7469
QY 283 aAspGlyMetGlnPheGluLeuSerProIleTyrHisValAlaAlaIleAspIlePhe 303
Db 7468 AGACGCGCATCAATTTGAACCAATCAACCATGTATCATGTGGAAGTACTCAAAAGCCCTTCT 7409
QY 303 uLysAlaTyrGlySerAlaLysArgValAsnLeuGluLysGluPheProGlnSerTyrVa 323
Db 7408 GGAATTA-----ACTGCTCTAGTCCCTGATTAATCTTCCACAGCTTAG 7367
QY 323 lGlnThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuProAspTyrAsnTh 343
Db 7366 ACCTACTTTGCTTGCATGAGTATTATCTCCTCAAAATGACTGGTCCAGATCATAAACA 7307
QY 343 rPromMetPheGlyAspSerTrpIleThrAspLysAsnPheArgMetAlaGlnPheAla 363
Db 7306 AATCCCTCTTGTGATAGCGATGTACAGATACCAAGAGATATCTTAACGCTTGCAGCCAC 7247
QY 363 rPheAlaThrAspGlyLysGlnGly-----TrpAlaArgValPheProAlaAsnGlnAlaIleLys 376
Db 7246 TATCCTTGAAGAACCACTTAAAGCTGCTCTTTTCCACACTTGAT--ATCGATAG 7190
QY 376 rPheAlaThrAspGlyLysGlnGly-----LysAl 386
Db 7189 TCTTCTCTTCTAGGAGAAAAGGTGTCACACTTTTGAACAGTTTACCCTCCAAACCTT 7130
QY 386 aProAsnPheLeuSerLysAlaLeuSerAsnAlaGlyPheTyrThrPheArgSerGlyTr 406
Db 7129 ACCAACTTTT---GCGCATCACTTTTGAACATTTCCGGGACATTTACCATC----- 7084
QY 406 pAspLysAsnAlaThrValMetValLeuLysAlaSerProGlyGluPheHisAlaG 426
Db 7083 -AATCAAGAAATATTATCTTTTTCNAANAATGGCCCATCGGTAGTTCCTATACCA 7025
QY 426 nProAspAsnGlyThrPheGluLeuPheIleLysGlyArgAsnPheThrProAspAlaG 446
Db 7024 TAGCGACCAAAATAGTCTCTGTCTATTATTAAGGACCAACCACTTTTCTGTGATCG 6965
QY 446 yValPheValTyrSerGlyAspGluAlaIleMetLysLeuArgAsnTrpTyrArgGln 466
Db 6964 GCGATACACTTACAAGAGAA-----CCATTACGCTATGCCCTTAAAGAGCGC 6917
QY 466 rArgIleHisSerThrLeuThrLeuAspAsnGlnAsnMetValIleThrLysAlaArg 486
Db 6916 TAGCCATCATCTACCGCTTCTTAGAAGAACAA-----TTACCTGAGCAGATTGA 6866
QY 486 nAsnLysTrpGluThrGlyAsnAsnLeuAspValLeuThrTyrThrAsnProSerTyr 505
Db 6865 TAGCTCTCTGGCT-----TACCTTTCTTACCCCAATCCAATTATTG 6824
QY 505 ----- 505
Db 6823 TCACCTTCGACAAAATGGACAGCTTTTACTTTATTGAGGTAGTACCACAGCAATCTTC 6764
QY 506 -----ProAsnLeuAspHisGlnArgSerValLeuPheIleAsnLysTyrPhe 523
Db 6763 TGATAGGAACAACCTACCAACATGACCGCAAAATCTCATCTTACCACAGGAATTTTCT 6704
QY 523 uValIleAspArgAlaIleGlyGluAlaThrGlyAsn----- 535
Db 6703 TATCATTTGATACGATA-----CAGCTCAAGGCAACCACTGCTTGTGTGAGTCAATTTAT 6650
QY 536 -LeuGlyValHisTrpGlnLeuLysGluAspSerAsnProValPheAspLysThrLysAs 555
Db 6649 TTTTGTATATCACTTAGACGTAAAGACAGAT----- 6619

```



```

QY 555 nArgValTyrThrTyrArgAspGlyAsnAsnLeuMetIleGlnSerLeuAsnAlaAs 575
Db 6618 -----CACCCTAGTACCTAGACTATCATGTA 6590
QY 575 pArgThrSerLeuAsnGluGluGlyLysValSerTyrValTyrAsnLysGluLeuLy 595
Db 6589 C---TGCCCTTTCACATACGAGAAACGATTGTGTCACAAAATAACAACCATACCTGAC 6533
QY 595 sArgProAlaPheValPheGluLysProLysLysAsnAlaGlyThrGlnAsnPheValSe 615
Db 6532 TAGTCACAACTCATCAAGCAAAACCTTTAAAGACAGGGTGTGCACCACT-----AC 6479
QY 615 rIleValTyrProTyrAspGlyGlnLysAlaProGluIleSerIleArgGluAsnLysG1 635
Db 6478 CCTATTAGTCCCTGACATACCAAGTACTCCTTAACACCACTTCAACACGCAACG 6419
QY 635 yAsnAspPheGluLysGlyLysLeuAsnLeuThrLeuThrIleAsnGlyLysGln 653
Db 6418 CAACCAATTCGACTGCC-----CTTCTCGCATTTAAAGAAAGCAAGCAG 6373

RESULT 5
AE010002/C
LOCUS Streptococcus pyogenes strain MGAS8232, section 50 of 173 of the
DEFINITION complete genome.
ACCESSION AE010002 AE009949
VERSION
KEYWORDS AE010002.1 GI:19747861
SOURCE Streptococcus pyogenes MGAS8232.
ORGANISM Streptococcus pyogenes MGAS8232
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
REFERENCE 1 (bases 1 to 12187)
AUTHORS Smoot,J.C., Barbian,K.D., Van Gompel,J.J., Smoot,L.M.,
Chaussee,M.S., Sylva,G.L., Sturdevant,D.E., Ricklefs,S.M.,
Porcella,S.F., Parkins,L.D., Beres,S.B., Campbell,D.S., Smith,T.M.,
Zhang,Q., Kapur,V., Daly,J.A., Veasy,L.G. and Musser,J.M.
Genome sequence and comparative microarray analysis of serotype M18
group A Streptococcus strains associated with acute rheumatic fever
outbreaks
Proc. Natl. Acad. Sci. U.S.A. 99 (7), 4668-4673 (2002)
MEDLINE 21927593
PUBMED 11917108
REFERENCE 2 (bases 1 to 12187)
AUTHORS Smoot,J.C., Barbian,K.D., Van Gompel,J.J., Smoot,L.M.,
Chaussee,M.S., Sylva,G.L., Sturdevant,D.E., Ricklefs,S.M.,
Porcella,S.F., Parkins,L.D., Beres,S.B., Campbell,D.S., Smith,T.M.,
Zhang,Q., Kapur,V., Daly,J.A., Veasy,L.G. and Musser,J.M.
Direct Submission
Submitted (31-JAN-2002) Laboratory of Human Bacterial
Pathogenesis/Rocky Mountain Laboratories/NIH, 903 S. 4th St.,
Hamilton, MT 59840, USA
FEATURES
source
1. 12187
/organism="Streptococcus pyogenes MGAS8232"
/strain="MGAS8232"
/db_xref="taxon:186103"
complement(106..339)
/gene="spyM18_0686"
complement(106..339)
/gene="spyM18_0686"
/note="best blastp match gb|AAK33591.1| (AE006517)
hypoetical protein [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="hypoetical protein"
/protein_id="AAL97359.1"
/db_xref="GI:19747862"
/translation="MEFVALFLSGSKSRVTSLPVLISLLPQQAAVLAVHTRRKK
SVVKSAEQPMLEHTGTGFSIURLKVELQSF"
complement(411..1451)
/gene="spyM18_0687"
complement(411..1451)

```

```

/gene="spyM18_0687"
/note="best blastp match gb|AAK33592.1| (AE006517)
conserved hypothetical protein [Streptococcus pyogenes M1
GAS]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL97360.1"
/db_xref="GI:19747863"
/translation="MTAALLHDIGHGAYSHTEALFHTDHEAFTQBIITNPETEINAI
LYRVAFDPDKVASVINHTYPNKVQLISQIDCDRMDYLLRDSYFSAANYGQFDLM
RILYRVPVEDGIVPEHSGMAVEDVIVSRFOMYVFPASRAVELILQNLKRAQ
HLYPQQAQKVTAPGLIPFPFKANLADYIALDDGWMNTYFQVWASEDHILSLDAS
RIRNKLKASVTFDQSRGELERQLVESVGDPDIYTGIIHINFDLPDIYRPELEN
PTQLIEMMQDGLSALQSPLVKALTGTTGYDRRFYFPKEMLELDDLFAFSKETFM
SYISNGHPHFQS"
1794..2180
/gene="spyM18_0688"
1794..2180
/gene="spyM18_0688"
/note="best blastp match gb|AAK33593.1| (AE006517)
hypoetical protein [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="hypoetical protein"
/protein_id="AAL97361.1"
/db_xref="GI:19747864"
/translation="MKQLQTNHIRFDETEIIQIHCWEKREKGGYQYLIYQNTDKK
VVIKYNELQATMSRFSNQSIIMKFPAGKVLIALPTPMGVQQLTDTSHYLDSCQK
LDLHVHLLQAQTEMLFASVHLELSD"
2410..5091
/gene="pacL"
/note="spyM18_0689"
2410..5091
/gene="pacL"
/note="best blastp match gb|AAK33594.1| (AE006517)
putative calcium-transporting ATPase [Streptococcus
pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative calcium-transporting ATPase"
/protein_id="AAL97362.1"
/db_xref="GI:19747865"
/translation="MSKQRHEAFYTSQSEETVLAQLETSREGLTSAQAKERIAEYGRN
EIDEGEKSLFMKFLDQFKDMLIILVAALSVLTEGEGLTDAIILAVILNAAF
GYQGOQAEAAIALKSMSPILARRDGHVTEIDSKELVPGDIIVLEAGDVVPADLR
LLEANSKTEEAALTGESVPVEKDLSTAVSEDAGIGDRVMGYQSNVTYGRGIVIT
NTGMYTEGHIAGLANADETTPKLNLDNLSKILTVAILVIAVTPAVGVFLRGQH
PLEGLMTSVALAVALPESGLPAIVTVVLSLTQVLAKNAIRKILPAVETLSLELLA
SDKTGLTNQMTVEKVTYINGLQSSADIAFDNTTLRVNPFANDTKVDFSGKSGDP
TETALVQGLDHNFDVRAMVAPRAELPFDSDRKIMSTIHKQADKYFIADVKGAPD
QLLRVQTQIBENGQIRPITDAKKTILTNTKSLAKQALRVLMVAKYSDALPTLEIE
VEANLVFGLVGMIDPERPEAAQVAKAEGIRPIMITGDHDTAKAKLRLGITEE
DGVDFVTGAELNELSDEEFQVFKQYSVYARVSPEHKVRIVKAQNEGKVVAMTGDG
VNDAPSLKATDIGICMGITGTEVSKGASMDVLADNFATIIIVAREGRKVFENQKTI
QYLLSANAEVFTIFALIFGMDVLQPVHLLINLVITDLPALVAGFAEPGVMVKKH
PGRKRSFFDGGVKEALYIQGAFQTLVLGVGFALMPPEPSTSYHDHADALTMAYVT
LGLIQLVHAYNVKYSQISFTVGLFPKLFNYSIEPAVFALMATVTVVFGNFQFFHVT
LTITQMLVVIIGSLMVLVVELVKAQVRSLSQDEKAI"
complement(5175..6170)
/gene="spyM18_0692"
complement(5175..6170)
/gene="spyM18_0692"
/note="best blastp match gb|AAK33595.1| (AE006517)
putative transcriptional regulator (LacI family)
[Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="AAL97363.1"
/db_xref="GI:19747866"
/translation="WQRVTKIDIAELAKTSKTTVSFYLNGRFDKMSSEETKNRIKSTI
EATNVKPSIARSINAKSKLIGVIGDITNSFNQIVKGIKSAQEGFYQLIIGNSN

```

```
YDPSREDELEKMLNLGVDFIIQIOTFNFRKYSRIIDIKKKKVVFFDSQLYEHRTNWV
KTNVADYVDIQQCIDKGYEHFIMITGNPNLLSTRIRASSGFDIVLEANHILTHEQMI
IDENQTSSEAIQAQFLOGSLTKSLVFNCHWALPKVFTAMKSLKFNIEPIGLVGFNDI
EWTFSSFTLITIIQPAVEGEQATKILIDIDEGHSQBAQOIIFDCQVNWQESTP"
/gene="spyM18_0693"
/complement(6234..8141)
/gene="spyM18_0693"
/complement(6234..8141)
/notes="best blastp match gb|AAK33596.1| (AE006517)
hypothetical protein [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative PTS dependent galactosamine IID
/protein_id="AAL97364.1"
/db_xref="GI:19747867"
/translation="MSLAPARFKETVPEDFCNKYLLDYQANSYADOKRIADLLNTNF
LFEDNWDMEPCSIPTLETTITQKAVTDDEPNFMNROTYLQTLILVYLKQBERYL
LAKGFIINWESAIPLDPKGLATRLDTGFCFAWVKRLIYLNLFYTKQESLIL
ASLEQLOFLHNTYLDKYSLSNWGILQTTAILADAYVGSIDIDIAAFAFAKELTQ
IALEQLEDSQPEQSTMYHVEVLSLLELTALVPDLPQLRPTLLAMSDYLLKMTDPD
HKQIPLGSDPTDTRDILTAAITILEEPHLKAAAPPTLIDISLILGKGVHTPEOLP
IQTLTPAHHEHSHITINQENYILFPKNGPIGSSHTSDQNSICLYKQGPLFCDA
GRYTYKEPLRYALKASHHSHTAFLEEQLPQIDSSWAYLSYPSKNSYCHLRQNGHYF
TEGYSQTFQSKNNYQHDRLILPPIGLFIIDITIAQCNHCLQSQFILDNHELVKTD
HLSDLRLSDCFPIETIETILSKKNQVILTSKLIKRPFKDKGCTSTLLVPDPTKVP
LTPLOTGRNPIETALSWMHLKGQFDSYICVLQEDLIKEXLIVLLNGHKIRGVVIVN
HITNELIRLKH"
/complement(8227..9048)
/gene="spyM18_0694"
/complement(8227..9048)
/gene="spyM18_0694"
/notes="best blastp match gb|AAK33597.1| (AE006517)
putative PTS dependent galactosamine IID component
[Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative PTS dependent galactosamine IID
component"
/protein_id="AAL97365.1"
/db_xref="GI:19747868"
/translation="MTGSKNFKKDYLLKTLARFPLQNGFNYSNVOGIGYANVYIPAL
KKHFGSKGLYQALDNCDFYNPHLPFTLSLHVLNDRPEEETRNIKMLMG
PLAGTGDLSQFCPLAFSTAAASLASDGLVIGTFFFLAMNIIITAKIGSGLYGK
VGTSPIDKLSQMAIVSRMANIVGTVIAGLAATSVKITVPTVPAAGKVDANAAQKE
VTIQMDLKIAPALLPALTLLMYLILKKNKTTYKLIVLITVIGVIGSWLGILA"
/complement(9035..9817)
/gene="spyM18_0695"
/complement(9035..9817)
/gene="spyM18_0695"
/notes="best blastp match gb|AAK33598.1| (AE006517)
putative PTS dependent N-acetyl-galactosamine-IIC
component [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative PTS dependent
N-acetyl-galactosamine-IIC component"
/protein_id="AAL97366.1"
/db_xref="GI:19747869"
/translation="WVNLILQALLGLWTAPCFSGMLGIYTNRCIILISFGVGIILGD
LPTALSMGAISELAYFGVAGGTVPNPIGPFIPTGLMIAITSAGKVTPEAALST
PIAIAIQLOTFFAYTAFAGATKQKQNGIRGFKFAANGTIWAFAFIIGLGLLG
ALSMPTLHLVDYIPVLLNGLTVAGKMLPAIGFAMILSVMAKKELIPFVLIGYVCAA
YLOIPTIGAIIGIIPALNEFVNKPKQVDATTVOGQDDWI"
/complement(9836..10324)
/gene="spyM18_0696"
/complement(9836..10324)
/gene="spyM18_0696"
/notes="best blastp match gb|AAK33599.1| (AE006517)
putative PTS dependent N-acetyl-galactosamine-IIB
component [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative PTS dependent
```

```
N-acetyl-galactosamine-IIB component"
/protein_id="AAL97367.1"
/db_xref="GI:19747870"
```

Alignment Scores:

Pred. No.:	5,85e-09	Length:	12187
Score:	233.50	Matches:	151
Percent Similarity:	36.12%	Conservative:	87
Best Local Similarity:	22.91%	Mismatches:	283
Query Match:	6.68%	Indels:	140
DB:	1	Gaps:	26

US-09-802-285a-2 (1-659) x AE010002 (1-12187)

QY	64	LeuAlaTyrTyrArgGluLysSerLysAlaArgGluProAspPheSer	-----	79
DB	8135	CTTGCTTTTCGACCGCTTTAAAGAACAGTC	---AATCAGACTTTTTCGACGAATTACCTG	8079
QY	80	-----AsnAlaGluLysProAlaAspIleArgGlnProIleAspLysValThrArg	96	
DB	8078	CTAGACTATCAGGCTAACAGCTATGCTGAT	-----CAAAAA	8043
QY	97	GluMetAlaAspLysAlaLeuValHisGln	-----PheGlnProHisLysGlyTyrGly	114
DB	8042	AGAATAGCTGACCTACTCTTAACCAATACCTTTTATTATTGAAGATAATTG	---GGATATGGA	7984
QY	115	TyrPheAspTyrGly	-----LysAsp	131
DB	7983	GCCTGCTCTATTCCTTATCGTTTGGAGACTATCACTTGGCNAAGCGCGTGACAGATGA	7924	
QY	131	pAsnGluValArgtrpGlnLeuHisArgValLysTrpGlnAlaMetAlaLeuVal	151	
DB	7923	TCCTGAGTGGAACTTTATGTTAAATCGACAGACCTACCTCCAAACATTAATCTCGTCTA	7864	
QY	151	RHisAlaThrGlyAspGluLysTyr	-----AlaArgGluTrpValTyrGlnTyrSe	168
DB	7863	CCTTGTGAAAGAGATGAACGGTATCTCTCTAAAGCCAAAGGTTCATCCTCAATGGAT	7804	
QY	168	rAspTrpAla	---ArgLysAsnProLeuGlyLeuSerGlnAspAsnAspLysPheVal	187
DB	7803	TGAATCAGCCATTCCTTTAGACCCCTAAGGGTTAGCTACTAG	---AACCTTAGATAC	7748
QY	187	pArgProLeuGluValSerAspArgValGlnSerLeuPro	---ProThrPheSerLeuPheV	207
DB	7747	GCATCCGTTGTTTCTGCTGGGTCAAATGTTTGATTACCTTAATCTTTTAAACGCTTAA	7688	
QY	207	alAsnSerProAlaPheThrProAlaPheLeuMetGluPheLeuAsnSerTyrHisGln	227	
DB	7687	CCAAGCAAGAGAAAGCCCTC	-----ATCCTAGCTTCCCTGGAAAAAC	7646
QY	227	InAlaAspTyrLeuSerThrHisTyrAlaGluGln	-----GlyAsnHisArgLeuP	244
DB	7645	AACCTCAATTTCTCATACCACTACTCTTGATAAATACAGCCTCAGCAATTGGGAACT	7586	
QY	244	heGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPheLysAspSerProA	264	
DB	7585	TGCAAAACCACCTCGCATTTTATTAGCTGATCGCTACTATGTTCTGACCTTGATATTGCTG	7526	
QY	264	rgTrpArgGlnThrGlyIleSerValLeuAsnThrGluIleLysLysGlnValTyrAlaA	284	
DB	7525	CAGCAACAGCTTTCTCTAGAAAAAGAAATTCAGCCAAATTCACATACAAATTCGCAAG	7466	
QY	284	spGlyMetGlnPheGluLeuSerProIleTyrHisValAlaAlaIleAspIlePheLeu	304	
DB	7465	ACGGTAGTCAATTTGAAACAATCAACCATGTCATCTGGAGGTGTATAATCCCTTTTGG	7406	
QY	304	ysAlaTyrGlySerAlaLysArgValAsnLeuGluLysGluPheProGlnSerTyrValG	324	
DB	7405	AATTA	-----ACTGCTCTAGTCCCTGATTATCTTCACACAGCTTAGAC	7364
QY	324	InThrValGluAsnMetIleMetAlaLeuSerIleSerLeuProAspTyrAsnThrP	344	
DB	7363	CTACTTTGCTTCCCATGAGTGTATTCTCTCAAAATGACTGATCTGACCAACACAGA	7304	

```

RESULT 6
SPNEU1908
LOCUS
DEFINITION
  Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***,
  in ordered pieces.
ACCESSION
  AL449930
VERSION
  AL449930.1 GI:11545155
KEYWORDS
  HTG; HTGS_PHASE2.
SOURCE
  Streptococcus pneumoniae.
  Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
  Streptococcus.
REFERENCE
  1 (bases 1 to 175936)
  Dopazo, J., Mendoza, A., Herrero, J., Caldara, F., Humbert, Y.,
  Friedli, L., Guerrier, M., Grand-Schenk, E., Gandin, C., de
  Francesco, M., Polissi, A., Buell, G., Feger, G., Garcia, E., Peitsch, M.
  and Garcia-Bustos, J.F.
  Annotated draft genomic sequence from a Streptococcus pneumoniae
  type 19F clinical isolate
  Microb. Drug Resist. 7 (2), 99-125 (2001)
  21335329
  11442348
REFERENCE
  2 (bases 1 to 175936)
  Dopazo, J., Mendoza, A., Herrero, J., Caldara, F., Polissi, A.,
  Humbert, Y., Friedli, L., Guerrier, M., Grand-Schenk, E., Gandin, C., de
  Francesco, M., Buell, G., Feger, G., Garcia, E., Peitsch, M. and
  Garcia-Bustos, J.F.
  Direct Submission
  Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A.,
  Severo Ochoa 2, 28760 Tres Cantos, SPAIN
  * NOTE: This is a 'working draft' sequence.
  * This sequence will be replaced
  * by the finished sequence as soon as it is available and
  * the accession number will be preserved.
FEATURES
  source
  1. .175936
     /organism="Streptococcus pneumoniae"
     /serotype="19F"
     /db_xref="taxon:1313"
     /clone="G54"
BASE COUNT 52994 a 31655 c 38793 g 52447 t 47 others
ORIGIN

Alignment Scores:
Pred. No.: 0.000179 Length: 175936
Score: 198.00 Matches: 140
Percent Similarity: 35.54% Conservative: 96
Best Local Similarity: 21.08% Mismatches: 273
Query Match: 5.67% Indels: 156
DB: 2 Gaps: 25

US-09-802-285A-2 (1-659) x SPNEU1908 (1-175936)
QY 71 SerLysAlaArgGluProAspPheSerAsnAlaGluLysProAlaAspIleArgGlnPro 90
DB 143534 TCAAGAAGCTATGAGAAAGTCAGAGTTCCTTGAACGGTTGATGACAAATCGTTTATG 143593
QY 91 IleAspLysValThrArgGluMetAlaAspLysAlaLeuValHisGlnPheGlnProHis 110
DB 143594 TTTGATAGT--CCTTGGGATATGGAGCCTTGTTCAAAATCCATCAATCCAGCCG--- 143647
QY 111 LysGlyTyrGlyTyrPheAspTyrGlyLysAspIleAsnTrpGlnMetTrpProValLys 130
DB 143648 -----ATGGTATGGGATCAAGTATTTGAAGAT 143674
QY 131 AspAsnGluValArgTrpGlnLeuHisArgValLysTrpTrpGlnAlaMetAlaLeuVal 150
DB 143675 GATCCGAATGGCTTATATGCTCAATCGAAGAATATCTCTTGAGTTTATGATAGG 143734
QY 151 TyrHisAlaThrGlyAspGluLysTyrAlaArg-----GluTrp 163
DB 143735 TATGTGGTAGAGGAGATAAGGACTATATTCAAAGTGCAGTCTTCTTCTATTGATTCG 143794

```

```

SPNEU1908
LOCUS
DEFINITION
  Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***,
  in ordered pieces.
ACCESSION
  AL449930
VERSION
  AL449930.1 GI:11545155
KEYWORDS
  HTG; HTGS_PHASE2.
SOURCE
  Streptococcus pneumoniae.
  Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
  Streptococcus.
REFERENCE
  1 (bases 1 to 175936)
  Dopazo, J., Mendoza, A., Herrero, J., Caldara, F., Humbert, Y.,
  Friedli, L., Guerrier, M., Grand-Schenk, E., Gandin, C., de
  Francesco, M., Polissi, A., Buell, G., Feger, G., Garcia, E., Peitsch, M.
  and Garcia-Bustos, J.F.
  Annotated draft genomic sequence from a Streptococcus pneumoniae
  type 19F clinical isolate
  Microb. Drug Resist. 7 (2), 99-125 (2001)
  21335329
  11442348
REFERENCE
  2 (bases 1 to 175936)
  Dopazo, J., Mendoza, A., Herrero, J., Caldara, F., Polissi, A.,
  Humbert, Y., Friedli, L., Guerrier, M., Grand-Schenk, E., Gandin, C., de
  Francesco, M., Buell, G., Feger, G., Garcia, E., Peitsch, M. and
  Garcia-Bustos, J.F.
  Direct Submission
  Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A.,
  Severo Ochoa 2, 28760 Tres Cantos, SPAIN
  * NOTE: This is a 'working draft' sequence.
  * This sequence will be replaced
  * by the finished sequence as soon as it is available and
  * the accession number will be preserved.
FEATURES
  source
  1. .175936
     /organism="Streptococcus pneumoniae"
     /serotype="19F"
     /db_xref="taxon:1313"
     /clone="G54"
BASE COUNT 52994 a 31655 c 38793 g 52447 t 47 others
ORIGIN

```

```

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
  source
  1. .175936
     /organism="Streptococcus pneumoniae"
     /serotype="19F"
     /db_xref="taxon:1313"
     /clone="G54"
BASE COUNT 52994 a 31655 c 38793 g 52447 t 47 others
ORIGIN

```

```

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
  source
  1. .175936
     /organism="Streptococcus pneumoniae"
     /serotype="19F"
     /db_xref="taxon:1313"
     /clone="G54"
BASE COUNT 52994 a 31655 c 38793 g 52447 t 47 others
ORIGIN

```

```

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
  source
  1. .175936
     /organism="Streptococcus pneumoniae"
     /serotype="19F"
     /db_xref="taxon:1313"
     /clone="G54"
BASE COUNT 52994 a 31655 c 38793 g 52447 t 47 others
ORIGIN

```

QY	164	ValTyrGlnTyrSerAspTyrAlaArgLysAsnProLeuGlyLeuSerGlnAspAsn	183
Db	143795	ATGAGCAGGTGAGAGAAATTTCTCTCAATCTTGATGACTAGAACCTTGATCGGT	143854
QY	184	-----LysPheValTyrArgProLeuGluValSerAspArgValGlnSerLeuPro	201
Db	143855	ATTGCTTCTTACCTTGGTTGAACCTA	143881
QY	202	ThrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGluPhe	221
Db	143882	---CTCTTGCTCTCTGAAATTTGACTTTGTAGAGGAGAAAGAACTAGAGAAATTTTG	143938
QY	222	AsnSerTyrHisGlnGlnAlaAspTyrLeuSerThrHisTyrAlaGluGln	238
Db	143939	GTCAGTCTAGAAAGCAGATGACATTTATGAAAGCCATATTCGCCCAAGTACACCTT	143998
QY	239	GlyAsnHisArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGlu	258
Db	143999	AGTAACTGGGGATTTTACAAACAATTCGATGCTTGTCTATTATTCTTTTTCAGAT	144058
QY	259	PheLysAspSerProArgTyrArgGlnThrGlyIleSerValLeuAsnThrGluIleLys	278
Db	144059	ARGATGGACCTAGAGAAGCTTACCATTGCTTTCAGAGAGTTGAAACAGCAAAATTTGAG	144118
QY	279	LysGlnValTyrAlaAspGlyMetGlnPheGluLeuSerProIleTyrHisVal	296
Db	144119	ACACAGATTTTAGGAGATGAGTCACTTTTGAAACAGTCACTCTCATCATGTAGAGTT	144178
QY	297	-----AlaAlaIleAspIlePheLeuLysAlaTyrGlySerAlaLysArgValAsnLeu	314
Db	144179	TATAAGCCTTACTGGATTGTGTCTC	144208
QY	315	GluLysGluPheProGlnSerTyrValGlnThrValGluAsnMetIleMetAlaLeuIle	334
Db	144209	CTTCCAGATTCGACAGATAGTACCAGAGTTGCTGGAAAGATGGCGACCTATATTCAA	144268
QY	335	SerIleSerLeuProAspTyrAsnThrProMetPheGlyAspSerTyrIleThrAspLys	354
Db	144269	ATGATCACAGGCTTAGATGACGAACCTTTGGCTTTTGGTGATGCGATTCTACAGAAACG	144328
QY	355	AsnPheArgMetAlaGlnPheAla	362
Db	144329	ACAGAAATTTTGACGCTGTCTGCTGTTTGAACAAGAGACCTCTTAAACGGTCTG	144388
QY	363	-----SerTyrAlaArgValPheProAlaAsnGlnAlaIleLysTyrPhe	377
Db	144389	GATGTTAAAGTTGATTTGCTTGTAGCTCTTGTCTGGGGCGAGAAAGGTCAAGCGACTG	144448
QY	378	AlaThrAspGlyLysGlnGlyLysAlaProAsnPheLeuSerLysAlaLeuSerAsnAla	397
Db	144449	CAGGAATTTGAAAGAGAGCTTGGCAGCCTAAG-----TCCATGATCTTTGAAGACTCT	144502
QY	398	GlyPheTyrThrPheArgSerGlyTyrAspLysAsnAlaThrValMetValLeuLysAla	417
Db	144503	CGACATGCTCATTAAGGAT-----GAACATCGTTTATCTATTTTCAAAAAT	144550
QY	418	SerProGlyGluPheHisAlaGlnProAspAsnGlyThrPheGluLeuPheIleLys	437
Db	144551	GGTCCCTAGAGAGTCCCATGCCATAGTACGAGAAATGTTTTTTCCTTACGTATCAA	144610
QY	438	GlyArgAsnPheThrProAspAlaGlyValPheValTyrSerGlyAspGluAlaIleMet	457
Db	144611	GGCCAACCTATTTTCATAGATGCTGGCGTATTCTTATTCG-GGAGATATATGAACGTTA	144669
QY	458	LysLeuArgAsnTyrTyrArgGlnThrArgIleHisSerThrLeuThrLeuAspAsnGln	477
Db	144670	TCTCTTAAAGAGTGTGGAGTCATTCGACCTGCATTGT-----	144708
QY	478	AsnMetValIleThrLysAlaArgGlnAsnLysTyrTyrGlu-----ThrGlyAsnAsnLeuAsp	496
Db	144709	-----AGATGGAAAGCTCCCGAAGAAATCAGGG	144738
QY	497	ValLeu-ThrTyr-----	500

Db	144739	ATCCTGGGAATATGATACTATCTCTCACTCCCTGTTTGTCCACATAAAGAGGAGG	144798
QY	501	-----ThrAsnProSerTyrProAsnLeuAspIle	510
Db	144799	AGTGCATTATATTGAGGGGCTTATTGTCAGCAAGACTGATTTGGCTTATCTT---CA	144855
QY	510	sGlnArgSerValLeuPheIleAsnLysLysTyrPheLeuValIleAsp-----ArgAl	528
Db	144856	CAAGAGAAATTCCTCATGTTGTTAGAGATCTCGCTCTTGGTAGATGACATCAGGTG	144915
QY	528	alleGlyGluAlaThrGlyAsnLeuGlyValHisTyrGlnLeuLysGluAspSerAsnPr	548
Db	144916	TCAAGTCTAG-----CATGAGCGGTTGACTCAGTTT-----	144946
QY	548	oValPheAspLysThrLysAsnArgValTyrThrTyrArgAspGly-----AsnAs	566
Db	144947	ATCCTTGACAAGGAT-----GTGACCTATCAAGATGGGAAATCAATCA	144990
QY	566	nLeuMetIleGlnSerLeuAsnAlaAspArgThrSerLeuAsnGluGluGlyLysVa	586
Db	144991	GTGAGACTATCGAGT-----GAAGTTGATTTGATTTGGAGATACCATCAT	145038
QY	586	lSerTyrValTyrAsnLysGluLeuLysArgProAlaPheVal-----Ph	601
Db	145039	TTCTCTTAATACAT---GAGCTTGAAAGAGTAGCAAACTCACCAGCGCAATTTCTT	145095
QY	601	eGluLysProLysLysAsn-----AlaGlyThrGlnAsnPheValSerIle-----	616
Db	145096	TCAGAAATCAGATCGCTGGATTATACCATCATTCGCCGATGAGATTTTGAATCATCGTCA	145155
QY	617	----ValTyrProTyrAspGlyGlnLysAlaProGluIleSerIleArgGluAsnLysG	635
Db	145156	TTCTGTCTACAGACAGATCATCGTAGAGTGAATAATGCTTTGGCTTTTGAATGAATAA	145215
QY	635	yAsnAspPheGluLysGlyLysLeuAsnLeuThrLeuThrIleAsnGlyLysGlnIle	655
Db	145216	TCAGCAAAACAGACAAGCTGATTCTGTATTATTAAGCGAGGATATTCGTGTAGTGAATAAT	145275
QY	655	uValLeuVal	658
Db	145276	GTGCCTCGTT	145285

RESULT	7
LOCUS	AE008410
DEFINITION	Streptococcus pneumoniae R6 section 26 of 184 of the complete genome.
ACCESSION	AE008410
VERSION	AE008410.1
KEYWORDS	GI:15457839
SOURCE	Streptococcus pneumoniae R6.
ORGANISM	Streptococcus pneumoniae R6
REFERENCE	1 (bases 1 to 11333)
AUTHORS	Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszcak, L., Burgett, S., DeHoff, B.S., Estrem, S., Fritz, L., Fu, D.-J., Fuller, W., Geringer, C., Gilmour, R., Glass, J.S., Khoja, H., Kraft, A., Lagace, R., LeBlanc, D.J., Lee, L.N., Lefkowitz, E.J., Lu, J., Matsushima, P., McAhren, S., McHenry, M., McLeaster, K., Mundy, C., Nicase, T.I., Norris, P.H., O'Gara, M., Peery, R., Robertson, G.T., Rockey, P., Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G., Zook, C., Baltz, R.H., Jaskunas, S.R., Rostock, P.R. Jr., Skatrud, P.L. and Glass, J.I.
TITLE	Genome of the bacterium Streptococcus pneumoniae strain R6
JOURNAL	J. Bacteriol. 183 (19), 5709-5717 (2001)
MEDLINE	21429245
PUBMED	11544234
REFERENCE	2 (bases 1 to 11333)
AUTHORS	Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszcak, L., Burgett, S., DeHoff, B.S., Estrem, S., Fritz, L., Fu, D.-J., Fuller, W., Geringer, C., Gilmour, R., Glass, J.S., Hann, A., Khoja, H., Kraft, A., Lagace, R.,

LeBlanc,D.J., Lee,L.N., Lefkowitz,E.J., Lu,J., Matsushima,P.,
McAhren,S., McHenry,M., McLeaster,K., Mundy,C., Nicas,T.I.,
Norris,F.H., O'Gara,M., Peely,R., Robertson,G.F., Rocky,P.,
Sun,P.-M., Winkler,M.E., Yang,Y., Young-Bellido,M., Zhao,G.,
Zook,C., Baltz,R.H., Jaskunas,S.R., Rosteck,P.R. Jr., Skatrud,P.L.
and Glass,J.I.
Direct Submission
Submitted (27-JUL-2001) Infectious Diseases Research, Eli Lilly and
Company, Lilly Research Labs, Indianapolis, IN 46285-0438, USA
FEATURES
source
1. .11333
/organism="Streptococcus pneumoniae R6"
/strain="R6"
/db_xref="taxon:171101"
complement(448. .554)
/rpt_family="RUP element"
/rpt_type=dispersed
/rpt_unit=Rup A
complement(758. .1387)
/genes="kdgA"
/note="spr0287"
complement(758. .1387)
/genes="kdgA"
/EC number="4.1.3.16"
/codon_start=1
/transl_table=11
/label=spr0287
/product="2-keto-3-deoxy-6-phosphogluconate aldolase"
/protein_id="AAK99091.1"
/db_xref="GI:15457840"
/translation="MTKSDTIIEIKKQIVAVIRGNTKEGLQASACIKGGIKATEI
AYNQVAGGIIKELVDLYDDQSVICAGVLTAVTADAILAGANVVPSPHAEIA
KMCNLYSTVPCITLTITWALEAGSEIKLIPGSAISPATISVAKAPIQVSVWV
TGGVDLNNIPQFAAGADAVGICGELKUNLASQGNFDRKISEIAQQYVTLR"
complement(1397. .2398)
/genes="kdgK"
/note="spr0288"
complement(1397. .2398)
/genes="kdgK"
/EC number="2.7.1.45"
/codon_start=1
/transl_table=11
/label=spr0288
/product="2-keto-3-deoxygluconate kinase"
/protein_id="AAK99092.1"
/db_xref="GI:15457841"
/translation="MTKILLFGEPLIRISPLDATSIGDHVASITYPGSSEINIAACNLQ
ALGISTKVTALPANIGDRFLTLKQHQIDTSSICRLGDRIGLYLENGFGQSEV
FYDRKTSISQIRPNMLDMSLFQGISHPFGSITVAIGOEVRAILLLKKEAKRGI
VVSMDLNLKMTLSLEAKYEFKSPARTDYCGIDPLMDONLEMPRDSASLEEV
ENMRLLKAYGFKATFHTLRSSDEODKNVQYQALEERFEESVOLKTAVYQRIQSGD
ARTSGALYQLLHSSSKTITIDFVAVASATLKCILPGDHLSTATSISNLLANAQDIIR"
complement(2429. .3070)
/genes="spr0289"
complement(2429. .3070)
/genes="spr0289"
/codon_start=1
/transl_table=11
/label=spr0289
/product="Hypothetical protein"
/protein_id="AAK99093.1"
/db_xref="GI:15457842"
/translation="MKIALINENSQAKNHLIYDSLKEATDKDYQLFNYGRGEGE
SQTYYQNGLMAILNTKAVDFVVTGCTGVGAMALNLSFFGVCGLAVDPTDAYLY
SQINGNALISIPYAKGFGAELTKLMFERLFAEEGGGVPRRVIPEQRARILNE
VQKITHTMLTKTDQDFLKDXTISGKYFQYFFENCQDDEVAAYLKEVLAK"
complement(3089. .3904)
/genes="gno"
/note="spr0290"
complement(3089. .3904)
/genes="gno"
/EC number="1.1.1.69"
/codon_start=1

/transl_table=11
/label=spr0290
/product="5-keto-D-gluconate 5-reductase"
/protein_id="AAK99094.1"
/db_xref="GI:15457843"
/translation="MTNTSFSIEQSLKKGKIALITGASYGIGFAIKSYAEAGATIVP
NDINQDLVNGKIEAVREVIGIEAHGVYCDVTDDEGICAMVKQIEQVEVIDLLVYNAGI
IRRVPCMSAADFRKVIDIDLNAPHIVKAVIPSMIKKGHGKIINICSMWSEIGRET
ISVYAAKGGKMLTRNIASEYGGANICQNGIPGYIATPQTAPRLQEDDGRHPFD
QFTIAKTPAARGNPELGMGPAVFLASDASNFVNGHILYVDGGILAYIGKQPE"
4175. .4609
/genes="PTS-EII"
/note="spr0291"
4175. .4609
/genes="PTS-EII"
/note="Putative"
/codon_start=1
/transl_table=11
/label=spr0291
/product="Phosphotransferase system sugar-specific EII
component"
/protein_id="AAK99095.1"
/db_xref="GI:15457844"
/translation="MKIVLVGHGHPATGIYSSLIQAGNENVEAIDFVEGMSADELK
OKILLAISNEEEVLILSDLLGSGPPKVSSTIMGENAKTMVLSGLNLAIMEAVFAR
MAHSPDEVVNVKSVAAQGGVNGKELFSDAEEEEEDFESGI"
4621. .5811
/genes="ugl"
/note="spr0292"
4621. .5811
/genes="ugl"
/codon_start=1
/transl_table=11
/label=spr0292
/product="Unsaturated glucuronyl hydrolase"
/protein_id="AAK99096.1"
/db_xref="GI:15457845"
/translation="MKIKVTIEKIKSPERFLEVLPLTKEEVQAIDKVIQLEINLDY
FKEDFPTATFDNVYPIMDNTIEWTGFTELAYEYSQQDAFKNAHKNVLSFLDR
VKNRVELDHDLDGLFYTPSCMAEYKINGDEAREATLKAADKLIERIQEGGFQAWG
DLGKKEHYRLIIDCLNIQLLPFAOETGQDYDIAESHFYASANNVIRDDASSPHT
FYFDPETGQPFKGVTRQGYSDSDSWARGOSGWYVGLTYRHLKDSGSCFDLPKGVTVY
FANRLPKDHSVWDLIENDGSDQSDSSATAVACGIHEMLKHLPEVDADKDIYKHAM
HAMLASLIEHYANDQFTPGTSLHLHGVIYSHSGKGVDEGNIMWDIYILEALIRFYKDM
NLYW"
5822. .6313
/genes="PTS-EII"
/note="spr0293"
5822. .6313
/genes="PTS-EII"
/note="Putative"
/codon_start=1
/transl_table=11
/label=spr0293
/product="Phosphotransferase system sugar-specific EII
component"
/protein_id="AAK99097.1"
/db_xref="GI:15457846"
/translation="MTMPNIIMTRIDERLIHQGQWLWKYLCGNTVIVANDEVSTDKM
QQTLMKTVVDSVAMRFFLPQKVIDIIHKANPAQTFFIVKDKDALTLVEGVPRIKE
INIGNIHAPGKEQVTRSFILGEEDKAALKELSQTHQVTFNTKTPTTNGDGAQVQNM
DYI"
6328. .7107
/genes="PTS-EII"
/note="spr0294"
6328. .7107
/genes="PTS-EII"
/note="Putative"
/codon_start=1
/transl_table=11
/label=spr0294
/product="Phosphotransferase system sugar-specific EII
component"

```

/protein id="AAK99098.1"
/db_xref="GI:15457847"
/translation="MSINVFQAILLGLWTARCFSGMLLGIYVNRGIVLSPFGVILIGD
LPTALAMGALGELAYMGVGAGGTVPNPPIGPIGFTLMAITSAGKVSPPAAALST
PIAVIQLOTQFAYTVRAGAPETAKHNLKFKFTLNATLWLFPIGFTGLGCLG
ALSDMTLLKLVYIPVLLTGLTVAGKMLPAIGFAMILSVMAKKELIPFVLLGYVCAA
YLNIPITGIAVGTIFALIEFYFNKPKTADHVVEBAHDDWI"
7094..7912
/gene="PTS-EII"
/feature="PTS-EII"
/feature="spr0295"
7094..7912
/gene="PTS-EII"
/feature="Putative"
/codon_start=1
/transl_table=11
/label=spr0295
/product="Phosphotransferase system sugar-specific EII
component"
/protein id="AAK99099.1"
/db_xref="GI:15457848"
/translation="MTGSNKLTKRDYLTSLRAFCQNGPNYSNYQGLGVANVMYPAL
KKGEDGGQFOALENCEFTNTNPHELFPITSLHLVLENGRPTKETRSIKMALNG
PLAGIDSLQFCLAPLESTIAASFAQGLVVGPIPLFLAMNTILITAKLSGLGYK
LGTVIDKLSQMATISRIANIIGTVIAGLAATSVKIMVPTFAAGBVKADAKOSIV
SIQGMKDVPALLPALFTLLVYLIKEKKWTKYKLVILTIVLIGISWUKIIA"
7912..8205
gene

Alignment Scores:
Pred. No.: 2,61e-05 Length: 11333
Score: 189.00 Matches: 139
Percent Similarity: 36.21% Mismatches: 100
Best Local Similarity: 21.06% Mismatches: 273
Query Match: 5.41% Indels: 148
DB: 1 Gaps: 28

US-09-802-285A-2 (1-659) x AB008410 (1-11333)
QY 71 SerLysAlaArgGluProAspPheSerAsnAlaGluLysProAlaAspIleArgGlnPro 90
DB 8302 TCARAAGACTAGGAAGTCAAGAGTCCCTTGACCGTTGATGACAAATCGTTTATG 8361
QY 91 IleAspLysValThrArgGluMetAlaAspLysAlaLeuValHisGlnPheGlnProHis 110
DB 8362 TTTGATAGT---CCTTGGATATGGAGCCTTGTTCAAAAATCCATCAATCCAGCGC--- 8415
QY 111 LysGlyTyrGlyTyrPheAspTyrGlyLysAspIleAsnTrpGlnMetTrpProValLys 130
DB 8416 -----ATGGTATGGATCAAGTATTTGAAGAT 8442
QY 131 AspAsnGluValArgTrpGlnLeuHisArgValLysTrpTrpGlnAlaMetAlaLeuVal 150
DB 8443 GATCCAGATGCTTATATGCTCAATCGACAGATATCTCTGCAGTTTATGATAGG 8502
QY 151 TyrHisAlaThrGlyAspGluLysTyrAlaArg-----GluTrp 163
DB 8503 TATCTGGTAGAAGGAGATAAGGCTATATCCAAAGTCGAAGTCTTTCTTATTTGATTGG 8562
QY 164 ValTyrGlnTyrSerAspTrpAlaArgLysAsnProLeuGlyLeuSerGlnAsnAsp 183
DB 8563 ATTGACAGGTGAGAGAAATTTCTCCTCAATCCTTGATGACTAGAACCTTGATAGCGGT 8622
QY 184 -----LysPheValTrpArgProLeuGluValSerAspArgValGlnSerLeuProPro 201
DB 8623 ATTCGTTCTTTACTTTGGTTGAACATA----- 8649
QY 202 ThrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGluPheLeu 221
DB 8650 ---CTCTTGCTCTCTTGAATTTGACTTGTAGAGGAGAGAACTAGAGAAAATTTTG 8706
QY 222 AsnSerTyrHisGlnGlnAlaAspTyrLeuSerThrHisTyrAlaGluGln----- 238
DB 8707 GTCAGTCTAGAAAGCAGATGATGCTTTATGAAGCTACTATCGGCCCAAGTACACCCCT 8766

```

```

QY 239 GlyAsnHisArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGlu 258
DB 8767 AGTAACCTGGGAGTTTACAAACAATTCGGATGCTTGTATCTATCTATCTTTTCAGAT 8826
QY 259 PheLysAspSerProArgTrpArgGlnThrGlyIleSerValLeuAsnThrGluLeuLys 278
DB 8827 AAGATGGATCTAGAAAGAGCTTACCATTGCTTCAGAGGAGCTTGAACAGCAAAATGAG 8886
QY 279 LysGlnValTyrAlaAspGlyMetGlnPheGluLeuSerProIleTyrHisVal----- 296
DB 8887 ACACAGATTTTAGGAGATGGAAGCAGTTGACAGTGCATCTCTATCATGTAGAGTT 8946
QY 297 -----AlaAlaIleAspIlePheLeuLysAlaTyrGlySerAlaLysArgValAsnLeu 314
DB 8947 TATAAAGCCTTCTGATTTGTGCTC-----TTG 8976
QY 315 GluLysGluPheProGlnSerTyrValGlnThrValGluAsnMetIleMetAlaLeuIle 334
DB 8977 CTTCCAGACTTTCGAAGATAGTTTACCAAGAGTTGCTGGAAGAGATGCGCACCTATATCAA 9036
QY 335 SerIleSerLeuProAspTyrAsnThrProMetPheGlyAspSerTrpIleThrAspLys 354
DB 9037 ATGATGACAGGCTTAGATGCGGACTTTGGCTTTTGGTAGAGCAATCTACAGAACG 9096
QY 355 AsnPheArgMetAlaGlnPheAla----- 362
DB 9097 ACAGAAATGTTGAGCCTGCTGCTGTTTGAACAAGAGACCTTCTTAACGCTCTG 9156
QY 363 -----SerTrpAlaArgValPheProAlaAsnGlnAlaIleLysTyrPhe 377
DB 9157 CATGTTAAAGTTGATTTGCTTAGCATCTTGTTCCTGGGCGAGAAAGTCAAGCGACTG 9216
QY 378 AlaThrAspGlyLysGlnGlyLysAlaProAsnPheLeuSerLysAlaLeuSerAsnAla 397
DB 9217 CAGGAATTTGAAGGAGAGCTTGGCAGCCTAAG-----TCCATGATCTTTGAAGACTCT 9270
QY 398 GlyPheTyrThrPheArgSerGlyTrpAspLysAsnAlaThrValMetValLeuLysAla 417
DB 9271 CGACATGCTGCATTAAAGAT-----GAACATCGTTATCTATTTTTCAAAAT 9318
QY 418 SerProGlyGluPheHisAlaGlnProAsnGlyThrPheGluLeuPheIleLys 437
DB 9319 GGTCCGCTAGGAAGTGCCTATAGCCATAGTACGAGAAATAGTTTGTCTACGATATCAA 9378
QY 438 GlyArgAsnPheThrProAspAlaGlyValPheValTyrSerGlyAspGluAlaIleMet 457
DB 9379 GCCCAACCTATTTTCATAGATGCTGGCGTTATCTTAT----- 9417
QY 458 LysLeuArgAsnTrpTyrArgGlnThrArgIle-----HisSerThrLeu 472
DB 9418 -----CGGAGATATATGAACGTTATCTCTTAAAGAGTGTGGAGTCAATTCGACCTGC 9471
QY 473 ThrLeuAspAsnGlnAsnMetValIleThrLysAlaArgGln-----AsnLysTrp 489
DB 9472 ATTGTAGATGG-----AAAGCTCCGAAAGAAATCAGGGATCCTCG 9513
QY 490 Glu-----ThrGlyAsnAsnLeuAspValLeuThr 499
DB 9514 GAATATGATATATCTCCTCCTCTGTTTGTCCACCATAAAGAAAGGAGGAATGCAT 9573
QY 500 TyrThrAsnProSerTyr-----ProAsnLeuAsp-----HisGlnArgSer 513
DB 9574 TATATTGAAGGGGCTTATTTGGTCAGCAGAACCTGATTTGCCCTATCTTCAACAGAGAAA 9633
QY 514 ValLeuPheIleAsnLysLysTyrPheLeuValIleAsp-----ArgAlaIleGlyGlu 531
DB 9634 ATCCCTCATGTTGGTAGAGGATCTCGGCTCTTGGTAGAGTACATCAGGTCTCAAGTCTAG 9693
QY 532 AlaThrGlyAsnLeuGlyValHisTrpGlnLeuLysGluAspSerAsnProValPheAsp 551
DB 9694 -----CATGAGCGGTGACTCAGTTT-----ATCCTTGAC 9723
QY 552 LysThrLysAsnArgValTyrThrTyrArgAspGly-----AsnAsnLeuMetIle 569

```

```

Db      9724  AGGAT-----GACCTATCAGATGGGAATCAATCAGTTGACACTA 9768
QY      570  GlnSerLeuAsnAlaAspArgThrSerLeuAsnGluGluGlyLysValSerTyVal 589
Db      9769  TGGAGT-----GAAAGTTGATTTGATTGGAGATACCATCATTTCTCTAAA 9816
QY      590  TyrAsnLysGluLeuLysArgProAlaPheVal-----PheGluLysPro 604
Db      9817  TACAAAT---GAGCTTGAAGAAGTAGCAAACTCACCAGCGCAATCTTTGAGAAATCAG 9873
QY      605  LysLysAsn-----AlaGlyThrGlnAsnPheValSerIle-----ValTyr 618
Db      9874  ATGCTGGATATACCATTCATGGCATGAGAGTTTGAATCATCGTCATTCGTCTAC 9933
QY      619  ProTyAspGlyGlnLysAlaProGluIleSerIleArgGluAsnLysGlyAsnAspPhe 638
Db      9934  CAGACAGATGATCGTGAAGTGGAAATGCTCTGGCTTTTGAAGTGAATAATGACGAACA 9993
QY      639  GluLysGlyLysLeuAsnLeuThrLeuThrIleAsnGlyGlnGlnLeuVal 658
Db      9994  GACAACTGATTCGTATTAAAGGAGGATATTCGTAGGTGAAAATTTGCGCTCGTT 10053

RESULT 8
SCF62
LOCUS       SCF62               34496 bp      DNA      linear      BCT 12-MAY-2002
DEFINITION Streptomyces coelicolor cosmid F62.
ACCESSION  AL121855 AL645882
VERSION    AL121855.3 GI:20520899
KEYWORDS   amidotransferase; aminotransferase; bi-domain oxidoreductase; blda
           codon; catalase; epimerase/dehydratase; glycosyl transferase;
           integral membrane protein; integral membrane transport protein;
           keta; lipoprotein; membrane protein; secreted protein;
           transcriptional regulator; transferase; UDP-glucose/GDP-mannose
           family dehydrogenase.
SOURCE     Streptomyces coelicolor A3(2).
           Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
           Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
           1 (bases 1 to 34496)
           Redenbach, M., Kieser, H.M., Denapate, D., Eichner, A., Cullum, J.,
           Kinashi, H. and Hopwood, D.A.
           A set of ordered cosmids and a detailed genetic and physical map
           for the 8 Mb Streptomyces coelicolor A3(2) chromosome
           Mol. Microbiol. 21 (1), 77-96 (1996)
           97000351
           8843436
REFERENCE  2 (bases 1 to 34496)
           Murphy, L. and Harris, D.
           Unpublished
JOURNAL    3 (bases 1 to 34496)
           Thomson, N.R., Parkhill, J., Barrell, B.G. and Rajandream, M.A.
           Direct Submission
           Submitted (29-OCT-1999) Streptomyces coelicolor sequencing project,
           Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
           CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
           David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
           Colney, Norwich, Norfolk NR4 7UH, UK
           On May 9, 2002 this sequence version replaced gi:6165435.
NOTES      Streptomyces coelicolor sequencing at The Sanger Centre is funded
           by the BBSRC and Beowulf Genomics
           Details of S. coelicolor sequencing at the Sanger Centre are
           available on the World Wide Web.
           (URL: http://www.sanger.ac.uk/Projects/s.coelicolor/) CDS are
           numbered using the following system eg SC7B7.01c. SC (S.
           coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
           strand).
           The more significant matches with motifs in the PROSITE database
           are also included but some of these may be fortuitous. The length
           in codons is given for each CDS.
           Usually the highest scoring match found by fasta -o is given for
           CDS which show significant similarity to other CDS in the database.

```

The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.gov/jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, ggg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid F62 Lies on the AseI-F genomic restriction fragment.

```

FEATURES             Location/Qualifiers
     source            1..34496
                        /organism="Streptomyces coelicolor A3(2)"
                        /strain="A3(2)"
                        /db_xref="taxon:100226"
                        /clone="cosmid F62"
                        complement(17..1123)
     misc_feature      /note="Pfam match to entry PF00083 sugar_tr, Sugar (and
                        other) transporter, score -87.50, E-value 0.00097"
     RBS               1315..1318
                        complement(1221..1226)
     RBS               1325..1327
     gene              /gene="SCF62.02"
                        /note="SCO0376"
     CDS               1325..1927
                        /gene="SCF62.02"
                        /note="SCF62.02, possible transcriptional regulator, len:
                        200 aa. Similar to several proteins of undefined function
                        from Streptomyces sp. e.g. Streptomyces coelicolor
                        TP:CAB51134 (EMBL; Y18817) hypothetical 20.6 KD protein
                        (190 aa), fasta scores opt: 352 z-score: 424.9 E():
                        2.6e-16 37.0% identity in 189 aa overlap. Contains a Pfam
                        match to entry PF01381 HTH 3, Helix-turn-helix between
                        residues 29..50 (+5.17 SD)."
                        /codon_start=1
                        /transl_table=11
                        /product="putative transcriptional regulator"
                        /protein_id="CAB58317.1"
                        /db_xref="GI:6066657"
                        /db_xref="SPTREMBL:Q9RJL2"
                        /translation="MLIMTQEDGDLSDLVKRIALRVAGWSLELAGRARLSQSSL
                        SRIENGRRLALDQVTLARALDTLDQLVETADEVITSPMIDSAHGLMRWPVKGDGP
                        GMTVVQRMTQPPENPARMRAPGRLWLVLSGTAVLMIGHRRFRVETHQAAEFPTM
                        MPHAITGAGGCPCEILGIFDRDARRGHQREGGEPEARDRAE"
                        1379..1543
                        /gene="SCF62.02"
                        /note="Pfam match to entry PF01381 HTH 3,
                        Helix-turn-helix, score 64.30, E-value 2.5e-15"
                        complement(2043..2576)
     gene              /gene="SCF62.03c"
                        /note="SCO0377"
                        complement(2043..2576)
     CDS               /gene="SCF62.03c"
                        /note="SCF62.03c, possible integral membrane protein, len:
                        177 aa. Contains multiple possible membrane spanning
                        hydrophobic domains."
                        /codon_start=1
                        /transl_table=11
                        /product="putative membrane protein"
                        /protein_id="CAB58318.1"
                        /db_xref="GI:6066658"
                        /db_xref="SPTREMBL:Q9RJL1"
                        /translation="MPEGAGRGSGVAPRGILRRRRPRLFKGIYGLVIALVALVAAL
                        DYPGGAANPGQDALVLLTALTSGAAHGVHIAQASGDGTAGPSRLKAVLAEPFLA
                        VAVLPVNMVLLAAEAGWAEETAADALTALVINTVLVLFALGTSAAFTAGLRLSCCRAGA

```

```

gene
  IDMLIGLVIIRADALIE"
  complement(2577..3311)
  /gene="SCF62_04c"
  /note="SCF62_04c"
  CDS
    complement(2577..3311)
    /gene="SCF62_04c"
    /note="SCF62_04c, possible secreted protein, len: 244 aa.
    Highly similar to Streptomyces coelicolor TR:CA852958
    (EMBL; AL109950) hypothetical 25.2 KD protein SC74.24C
    (242 aa), fasta scores opt: 1283 z-score: 1457.6 E(): 0
    83.1% identity in 242 aa overlap. Contains a possible
    N-terminal signal sequence."
    /codon_start=1
    /transl_table=11
    /product="putative conserved secreted protein"
    /protein_id="CA858319.1"
    /db_xref="GI:6066659"
    /db_xref="SPTREMBL:Q9RJL0"
    translation="VRMKIAVIGGTGLIGSQVVKOLNAGHQAVPHSRSTGVDVVGGG
    GVDDAVEAGAEVVNLTNSPTFDDASPAFFRASMDTLLAAARRAGTGHFVILSVGVDR
    VPALDYVRAKVLQEBILRDGPVPIVIRATQPMERMEAVLSWTASDDSVRLPATPIOP
    IAAKDVAARAVADVAGPLNGIRNIGVPEVFLDELGLTLIAHKGDARTVVDPAQM
    FGAVEGDLTDLDAHAPTRYADWLS"
    complement(3320..3324)
    RBS
      3580..3584
    RBS
      3591..5054
    gene
      /gene="SCO0379"
      /note="Kata"
      CDS
        3591..5054
        /gene="SCO0379"
        /note="SCF62_05, kata, catalase (EC 1.11.1.6) len: 487 aa.
        Has been previously sequenced and characterised from
        Streptomyces coelicolor strain ATCC10147 TR:P77948 (EMBL;
        X96981) catalase (EC 1.11.1.6) (488 aa), fasta scores opt:
        3300 z-score: 3768.6 E(): 0.99.2% identity in 488 aa
        overlap. Also similar to another proposed Streptomyces
        coelicolor catalase (EC 1.11.1.6) TR:Q9Z598 (EMBL;
        AL035478) SC2G5.25C (487 aa), fasta scores opt: 1704
        z-score: 1896.1 E(): 0.53.8% identity in 483 aa overlap.
        Contains Prosite hits to PS00437 Catalase proximal
        heme-ligand signature and to PS00438 Catalase proximal
        active site signature. Also contains a Pfam match to entry
        PF00199 catalase, Catalase."
        /codon_start=1
        /transl_table=11
        /product="catalase (EC 1.11.1.6)"
        /protein_id="CA858320.1"
        /db_xref="GI:6066660"
        /db_xref="SPTREMBL:Q9RJK9"
        /translation="MPENNQKPLTTVAGAPVDPNQNLSLTSGRPGMLLQDVMFLKLA
        HFDRVIERMHAAGSGAGFTVTVDITRYTSKIFSEIGKKTPLFRSTVAGER
        GAADARDIRGPAVKFYTDGNWDLVGNTPVFFRDLKPLPDINAHVKRPTNLRN
        AENWDFWNLPEALHOVTIYMSDRGIPASVYHMGFGSHYSLINABGRFWFKFHH
        RIQGIKNLTDAEALVKDREHQDLFDIEDGDFPKWLFYOVNPEADAENYRF
        HPFDLTWKKXQDYLIRVGEWELNPNPNFADYEAQFSPANVPVPGISFSPDRLQ
        GRLFSGDAQRYLGVNHHQIPVNAFPKNPVNSYHRDGMRVGNQCATPGVEPNYSYGR
        WQEQPAYRDPQAQVADARENYREDNDNYEPQGNLFQWSPQEQVLFEENTARAID
        GASATYIERHIGNCTQADPAYGAGYKALAEALAGNL"
        3618..5030
        misc_feature
          /gene="SCO0379"
          /note="Pfam match to entry PF00199 catalase, Catalase,
          score 1129.70, E-value 0"
          3726..3776
        misc_feature
          /gene="SCO0379"
          /note="PS00438 Catalase proximal active site signature"
          4596..4622
        misc_feature
          /gene="SCO0379"
          /note="PS00437 Catalase proximal heme-ligand signature"
          5090..5132
        stem_loop
          /note="Inverted repeat 21/21 (100%) with 1 base loop."
          5145..5148
        RBS
          5158..5553
        gene
          /gene="SCF62_06"
          /note="SCF62_06, unknown, len: 131 aa. Weakly similar to
          several prokaryotic Ank-like proteins e.g. Pseudomonas
          syringae TR:P95546 (EMBL; AF001355) AnkF precursor (183
          aa)
          Alignment Scores:
            Pred. No.: 0.000185 Length: 34496
            Score: 186.50 Matches: 105
            Percent Similarity: 33.33% Conservative: 61
            Best Local Similarity: 21.08% Mismatches: 198
            Query Match: 5.34% Indels: 135
            DB: 1 Gaps: 18
          US-09-802-285A-2 (1-659) x SCF62 (1-34496)
          QY 58 AspAlaAlaLysAlaLeuLeuAla-----Tyr 66
          Db 17535 GAGCGCGGAGAGCGCTCGTCCCGGAGCGGACCGCTGCTGCGCGGACGTCGAGTAC 17594
          QY 67 TyrArg---GluLysSerLysAlaArgLysProAspPheSerAsnAlaGluLysProAla 85
          Db 17595 TTCGGCGTGGCGCGGACGACCTGCTCGAGCGGACTGGTGC-----17636
          QY 86 AspIleArgGlnProIleAspLysValThrArgGluMetAlaAspLysAlaLeuValHis 105
          Db 17636 -----17636
          QY 106 GlnPheGlnProHisLysGly-----TyrGlyTyr---PheAspTyrGlyLys 120
          Db 17637 ---TTCGACCGGAGACCGGCGCGCGCGCGCGCTGCGGCTGACGCTCGACGTG-----17687
          QY 121 AspIleAsnTrpGlnMetTrpProValLysAspAsnGluVal-----134
          Db 17688 -----CCGTACCGCGACGAGGACGCGGCTCGGTGACATCAAGCAG 17726
          QY 135 ArgTrpGlnLeuHisArgValLysTrpTrpGlnAlaMetAlaLeuValTyrHisAlaThr 154
          Db 17727 ATCTGGAGCGCTCCCGGACGACGACGACGCTGCTGCGCGCGGCTGACGCGGTACAC 17786
          QY 155 GlyAspGluLysTyrAla-----ArgGluTrpValTyrGlnTyr 167
          Db 17787 GCGCAGCAGCGGTACGCGCGCGGTGCGCGGACGACCTGCGGTGCTGCGTGG-----17834
          QY 168 SerAspTrpAlaArgLysAsnProLeu-----GlyLeuSerGln 180
          Db 17835 -----TGGCGCGGCCAACGCGCGCTGCGCGGCTGCGGTGCGGTGCGGTGCGGTG 17888
          QY 181 AspAsnAspLysPheValTrpArgProLeuGluValSerAspArgValGlnSerLeuPro 200
          Db 17889 GGGATCCGGCTGCTCTCTCTG-----GTGTGGATCCCGCGGCTGCTGCGCGGTGCGCG 17942
          QY 201 ProThrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGluPhe 220
          Db 17943 GCGCGCGCGCGCTCTTCGAGGACAAACCGCGG-----CGCGTGGAGCAG 17987
          QY 221 LeuAsnSerTyrHisGlnAlaAspTyrLeuSerThrHisTyrAlaGluGlnGlyAsn 240
          Db 17988 ATCTGGCACCCACGAGCGTGGCTGCGCGCGCTTCCCGCGCGGCTGCTGCGCGCAACAC 18047
          QY 241 HisArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPheLys 260
          Db 18048 CACGTCTGGCGGAGCGCGCGGCGGAGTTGGCGCGGCGGCTGCGGCTTGGGTGGTCCCC 18107
          QY 261 AspSerProArgTrpArgGlnThrGlyLysSerValLeuAsnThrGluLysLysGln 280
          Db 18108 GAGTCGCGCAGCGTGGCGGAGCAGACGCGTGGCGTGGTGGTGGTGGTGGTGGTGGTGGTGG 18167
          QY 281 ValTyrAlaAspGlyMetGlnPheGluLeuSerProIleTyrHisValAlaAlaLeuAsp 300
          Db 18168 AGTTTCTCTCGCGCTCAACCGGAGCTGGCCACCGAGTACACCGGCTGGTGGTGGTGGTGG 18227

```



```

QY 301 IlePheLeuLysAlaTyrGlySerAlaLysArgValAsnLeuGluLysGluPheProGln 320
Db 18228 CTCGGCTGGCGGCTGGCGGAGCGGAGCGCGCGGTGTG-----CCGTGCCCCGG 18281
QY 321 SerTyrValGlnThrValGluAsnMetIleMetAlaLeuIleSerIleSerLeuProAsp 340
Db 18282 ACGTGGCGGCTGGTGTGGCGATGACGACGCGCTCGCGCGCTCGTGGACGACCGG 18341
QY 341 TyrAsnThrProMetPheGlyAspSer----- 349
Db 18342 CTGCGTCCGCGCGGACGCGGAGCGGACGACGCGCGCTGGTGGACGCGCGG 18401
QY 349 ----- 349
Db 18402 GGCACCGACCGCTGGGCTGGCTGCTGCCACCGGGGACCGGTGTCGGTGGCTTCCC 18461
QY 350 ---TTP-----IleThrAspLysAsnPheArgMetAlaGlnPheAlaSerTrpAlaArg 366
Db 18462 TGGTGGCGGACGCGTACCGCGGATGTCGACCGCGCTGCTGGCGGCTGGTGGTCCGG 18521
QY 367 ValPheProAlaAsnGlnAlaIleLysTyrPheAlaThrAspGlyLysGlnGlyLysAla 386
Db 18522 CCTACGCGGGGACGACGCG-----GGCGGCGCGCTGACCCCG 18560
QY 387 ProAsnPheLeuSerLysAlaLeuSerAsnAlaGlyPheTyrThrPheArgSer---Gly 405
Db 18561 CGCGCGCGGACGCGCGCCACTTCGCGGACGCGGCGCTACGGTGTGCGCGGCGCGGCG 18620
QY 406 -TrpAspLysAsnAlaThrValMetValLeuLysAlaSerProGlyGluPheHisAl 425
Db 18621 GGGATCTGGTGGCTGCGACGCGGCTGCGACGCGCTTCTGTCATCGCCGCG-CACGC 18679
QY 425 aglnProAspAsnGlyThrPheGluLeuPheIleLysGlyArgAsnPheThrProAspAl 445
Db 18680 CCACGCGGACGCGCTCTCGTGGAGTCCGCGACGCGGCGTCCGCTCGCCGACCC 18739
QY 445 aglyValPheValTyrSerGlyAspGluAlaIleMetLysLeuArgAsnTrpTyrArgl 465
Db 18740 GGGACCTACTCTACCGACGACGCGGTG-----TGGCGGGGTACTTCGGTCTC 18790
QY 465 nThrArgIleHisSerThrLeuThrLeuAspAsnGlnAsnMetValIleThr 482
Db 18791 GACGCTCGGCCACACACCTCGAGCTGACGCGCGCGACCGCTCGCTCC 18842

RESULT 9
AF498407 22913 bp DNA linear BCT 13-JUN-2002
DEFINITION Pseudomonas aeruginosa serotype 015 putative O-antigen biosynthesis
ACCESSION AF498407 AC104726
VERSION AF498407.1 GI:20559869
KEYWORDS
SOURCE Pseudomonas aeruginosa.
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
REFERENCE 1 (bases 1 to 22913)
AUTHORS Raymond,C.K., Sims,E.H., Kas,A., Spencer,D.H., Kutyavin,T.V.,
Ivey,R.G., Zhou,Y., Kaul,R., Clendenning,J.B. and Olson,M.V.
TITLE Genetic variation at the O-antigen biosynthetic locus in
Pseudomonas aeruginosa
JOURNAL J. Bacteriol. 184 (13), 3614-3622 (2002)
MEDLINE 22053227
PUBMED 12057956
REFERENCE 2 (bases 1 to 22913)
AUTHORS Raymond,C.K., Sims,E.H., Kas,A., Spencer,D.H., Kutyavin,T.V.,
Ivey,R.G., Zhou,Y., Kaul,R., Clendenning,J.B. and Olson,M.V.
TITLE Direct Submission
JOURNAL Submitted (04-APR-2002) Genome Center, University of Washington,
Box 352145, Seattle, WA 98105-2145, USA
COMMENT On May 13, 2002 this sequence version replaced gi:17975255.
FEATURES
Location/Qualifiers
1..22913
source

```

```

/organism="Pseudomonas aeruginosa"
/serotype="015"
/db_xref="taxon:287"
<1..104
/feature="ORF 1; C-terminal coding region of rpsa"
/codon_start=3
/transl_table=11
/product="rpsa"
/protein_id="AAM27639.1"
/db_xref="GI:20559870"
/translation="EKDAMKELRKQEVESAGPTTIGDLIRAQMNQOG"
241..525
/feature="ORF 2; himd/iHFB"
/codon_start=1
/transl_table=11
/product="Himd"
/protein_id="AAM27640.1"
/db_xref="GI:20559871"
/translation="MTKSELIERIVTHOGQLSAKDELAIKTMLSQMSOALATGDRIE
IRGSGFSLHYRAPVGRNPKTGESVRLDGKVFPHFKGKDRDRNEPE"
359..861
/feature="ORF 3; hypothetical orfa"
/codon_start=1
/transl_table=11
/protein_id="AAM27641.1"
/db_xref="GI:20559872"
/translation="MOWFRSSAAVAVLLVSVAVVVFLENQGVQLAFLGVQFPQWP
LAVYIIAIFILGGLGLAAQPLFTISRVRASSLRAELVQARKEVDTLRAEPSKVS"
1146..2174
/feature="ORF 4; wzz; similar to chain length determinant
protein"
/codon_start=1
/transl_table=11
/product="Wzz"
/protein_id="AAM27642.1"
/db_xref="GI:20559873"
/translation="MSKNEIDFELIQLWAQKVLVSLTFLGVLGASVAFLSKPV
YEARVAVPSSLSVAGNIARSRSGLPPTVSQVYSVFTNLSQEGTRQFFREVY
LPSLEERVVGLYRAFSDVSNQPTKELPERYTVVFDHPDPPQAKWAKQYIELA
AQRSMENRGAQLEYSRNIQQIDTLRQVVKERRDRIVRLKEALQVAEAVGVE
KPLATGSDIQQLSAIMEGSLMYMRGSRALRAEIALBARTSDDPFIPALRGLQOLA
LYESKQLSFDRAVPRQDSVETPDDPIKPRKVLILALGILAGVGLVLAALLRLER
EKRLDTG"
2421..3491
/pseudo
/codon_start=1
/transl_table=11
/3524..4786
/feature="ORF 6; similar to UDP-glucose/GDP-mannose
dehydrogenase"
/codon_start=1
/transl_table=11
/protein_id="AAM27643.1"
/db_xref="GI:20559874"
/translation="MSFOTISVIGTIGTLPATAVFAFRQKVKIVGVNQAVDTINR
GEIHVEPDLDMVHAATEGYLRATNPEDAPAFLIAVPTFFKHEVDPLTIFAAS
KSIAPVLKKGLVILESTSPVGAETQMWALAEFPDLFTFTHGEASDIRVAHCPER
VLPGHRELIVENDRVICGSEKSEASRLYKIFVGEICITINARTAMCKLTNSF
RDVNIAPANELSIIICDKLDINWELIRLANRHPRVNLPQPGVGCHCIAVDPTFVS
KTPDEARLRSAREVNDHKPFWLEKIKLAVADVLOKPEKTARDVSVACFLAFKAN
IDDLRESPAJIAAQVAVSHFGRVLAVEPNIEELPKFSGNLTLASTDSAMQEAADV
LLVHKFREIAPHPRQVVDITRGLWEA"
5040..6155
/feature="ORF 7; similar to Polysaccharide biosynthesis
protein; potential multiple membrane spanning domains."
/codon_start=1
/transl_table=11
/protein_id="AAM27644.1"
/db_xref="GI:20559875"
/translation="MLIHALVGYASIFDAGLGSVVRQIAISSAPEKRGPTLSSLL
FVLISGGLAAALVHFVFAEYLVVDIYFKVDRSSVEDALTGTVAALCIPFLASLVQY
LEGVEDFYSFNKRALSTLTLLPVAFLLPKFTFSSLIYGLPARAGSCFLIALIVS
RCSVWTRWDVKVFKTLISFGSWLITNIISIPVMVYMDKFLIARITAGELAYAAP

```

```
SELINRMSILPVAITRAVFPRLSALGOGSASVKRQALLYSGGLVLPITIIIVLFFVAPW
GMGLWPEYSGTAQVQIOMLVGLFNSIALVPVNNIQAQYSRTTAMVHLLVLPY
LMLWYFSVQOQGVGVAASISWLSRMAVDICILMAYLDSK"
6162..7412
/note="ORF_8; potential multiple membrane spanning
domains."
/codon_start=1
/transl_table=11
/protein_id="AAM27645.1"
/db_xref="GI:20559876"
/translation="MASYIENSLSKFLFFVLLTLMSPKSPFAVIKIVILAFMFSMWV
LHVLGGVRRSLDIFPYLAVIIAVLWGLNDGYALAGDAFLYVYVFSLFL
LHVLGGVRRSLDIFPYLAVIIAVLWGLNDGYALAGDAFLYVYVFSLFL
HLYWLVNRRSLDIFPYLAVIIAVLWGLNDGYALAGDAFLYVYVFSLFL
HEGFOLGNNVNSLFFTVPLLSALVCSSELSKRYELFFVSLVGLVIAVLLAGRA
LMLLILVPLAVPFRIRISGVSSIEGGRGVKLIPIVFFFAVGGVGLLAGLNVNQ
AERIESAPAGSVRDQFISLWKGTFDPLSGGSGVDVIRSEERPMWYELSTWQL
LPNGGLGFLSITICIGFFYLKALFVIKFNFLSYRAEALSTLVGLVFCVGSATNPYFS
GFDSLFLGFTPLIASVGKVAHS"
7399..8463
/note="ORF_9; similar to Glycosyl transferase"
/codon_start=1
/transl_table=11
/protein_id="AAM27646.1"
/db_xref="GI:20559877"
/translation="MRIREASMLKSNRVDPIIIVLLNNGWQDTIECLSLRHLEQPO
DILVFDNGSENDVYQLREYLRSISIAQHEEAFELDHVLHVIHYDVSGIVVRLCSLG
ENLFGACSNLAKVKAESMYVYALLANDTVLESDSLRIHMEVPSVDSADIVIPQIR
YFDRPILNCGGEVSRRGRVRYFSKGNLYLEKEAFVPGFATGCCMLPKVDFPRS
IGGTFQFFGREDVELSFRLLRRGTHVVCAARAIYHKVGASINGKDELIFRKTFFH
YLNRFNNRTQLPVAWNAWRLPIVFKVAITLRKNFGKSGSEILAFIRLLMRQSRRLP
GVDAKLFKSVLSKGLDYNDG"
8466..9469
/note="ORF_10; similar to Glycosyl transferases group 1"
/codon_start=1
/transl_table=11
/protein_id="AAM27647.1"
/db_xref="GI:20559878"
/translation="MVNAGRTIIVIGFPDPGVGTAFVSRILAENNMVSEVIDIYPS
SKSVPSFKGRVWPKFGTGLSYLANHROWAGRLHNFSTSLGAIYVFLVLKPFAS
EFSMLHSGSLQVKSRLYRALSLRDYVFCWNEGRQYRGVGLGVPEEKRLVSSY
IKIPLLSIVPRKDIQDFDGRSVLVASGFPSSDYNEHCIDVSVAREGLKLAIFLY
GSGDALNRIBERASERVKIPWNCQATFNHALSRASVILRPTSKDSFGVAVADAVEL
GVPVAVDCKRYPGAGLPVDDFSAFCTVDCILNGTDIRTEDESGFRAPFSYSLDL
RDV"
9466..11607
/note="ORF_11; similar to Zinc-binding dehydrogenases"
/codon_start=1
/transl_table=11
/protein_id="AAM27648.1"
/db_xref="GI:20559879"
/translation="MKVOLQDMAKGGTLVTEAPAPQASKEHIVINSRVSLISAGTERM
LVFGKASYLKEARQOQPKVWLEKVTQDGLMTIEAVRSKLGPGLPGYCNVGVVS
EVGSGVEGFKVGRVNSNGPHADVVKVSKNLCARIPDEVDDESAFVVAASIGQIR
LAQPTLGEAFVVTGVLGLTQVLLRAHGRVLAIDYDENKLSIARGYGAATNPGR
DYEKELTAMAFSRGGGVDTITATQSSDFVTHAARSRKRGRVILVGLTGELNRA
DFEYKELTQVSCSPGRYPDNPVEEKQDYPVLRWTEORNEFALIMLACGLDLY
OALLIHTREAFADAPQALVDKREKAGLILQVTAADAASRAIROLELROGVTFDPV
RVLGFIGAGNVASMLIPAKAAGAOFTIATAGINGVINGOKAGFAEASTVDAM
LANPINTVAVVTRHDSHARVSVQVLRAGKSVFVEKPLAIHAGELQEVDRDAYQANSN
GASQALMGFNRFRSPQVKMSLLAAVKEPKSFIMTNWAGAI PANHWTQDKAVGGGR
IIGEACHFIDMLRFLAGSRIVSVQASMGDGVETEDKASITILGPDGSGFTILYL
ANGAASFPERVEFTAGRLQDLONFRFKYGVWGFPSKLNLMKQDKQACAKAFLD
GLQSGQQAIPABEIPFVADVITQVAELLRNQ"
11604..14654
/psudo
/codon_start=1
/transl_table=11
/insertion_seq="undetermined"
11888..13089
14651..15868
/note="ORF_13; similar to Glycosyl transferases group 1"
/codon_start=1
/transl_table=11
```

```
/protein_id="AAM27649.1"
/db_xref="GI:20559880"
/translation="MRVIYFQHSTPKMGVIRSYEMARLVARGLVARGHOVTWVCSYGG
QGTGLDVPFSKGRQGTVDGDIIEFDLAYNSDTGFIKRITITFFSFAUKSRILAMTEE
YDLVFAITPTLITAGIPGIFARWLKAFVFEVRLMPELPMKAMGVITNPNVGLMMNSVL
EWLSYRSARHLIGLSPGIVEIGRLGPAHIVEMVPGNCDLIDIFASDEPWRPERVKA
DOLLAVTAGTHGVANGLDVLANAAELKKGRDDIKLLLVGQGLKPLVLOQRPKRGL
DNVVFPEPKNKARLAGLMAASADKGLQVLANPFAVYCTSPNKFEDYIAAGLPVNNYP
GWLAMTIADNOCGFVAFEPDSTAFADALERAADRESLRAMGERGRALAVREFRHL
ADRFDVMLEGAKS"
16021..16464
/note="ORF_14; similar to Bacterial sugar transferase"
/codon_start=1
/transl_table=11
/protein_id="AAM27650.1"
/db_xref="GI:20559881"
/translation="MIKPTNTRDAVDANGNALPDVMTMTPFCNFRSTSLDELPELWN
VLKGMDSLVGPRLLMEYLPYVSEQYRHLRPGVTGWAQVGNRNLMSWBEKFELDV
WYVDNRSPWLDKILMLTKKVFVREGISADGEATMSKFTGSEKS"
16461..17096
/note="ORF_15; similar to Bacterial transferase
hexapeptide (four repeats)"
/codon_start=1
/transl_table=11
/protein_id="AAM27651.1"
/db_xref="GI:20559882"
/translation="MRKLAIIAGGHHKVVADTAECGCEWEEVFPDDAWPGKSANGRW"

Alignment Scores:
Pred. No.:      0.000598      Length:      22913
Score:          177.50      Matches:      133
Percent Similarity: 36.52%      Conservative: 81
Best Local Similarity: 22.70%      Mismatches: 250
Query Match:      5.08%      Indels:      124
DB:              1          Gaps:      31

US-09-802-285A-2 (1-659) x AF498407 (1-22913)
QY 112 GlyTyrGlyTyrPheAspTyrGlyLysAspIleAenTrrpGlnMetTrrpValIysAsp 131
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 13077 GGGGAAGCTTACGGAATGACCAATCTCATCGAATTGG-----TGGGAATGCCCGAT 13130
QY 132 AsnGluValArg-----TrrpGlnLeuHisArgValIysTrrp 144
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 13131 TTCGATCTGGCAGTCGGAGACATCAAGCGTCTCTGGAGGCGCTCGCGTTTCGACTGG--- 13187
QY 145 GlnAlaMetAlaLeuValTyrHisAlaThrGlyAspGluLysTyrAlaArgGlu----- 162
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 13188 ---GTGCCTGCTTTGGCGCAAAAAGCTCTGGCTGTGAGAGTGTGGCTGGACAAACTG 13244
QY 163 -----TrrpValTyrGlnTyrSerAspTrrpAlaArgLysAsnProLeuGlyLeuSerGln 180
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 13245 AATGCCTGGTGG-----AACGACTGGTGCATGAATAATCTCCTTACAAAGGGCGG 13295
QY 181 AspAsnAspLysPheValTrrpArg---ProLeuGluValSerAspArgValGlnSerLeu 199
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 13296 AAT-----TGGAAATCGCGCCAGGAGGCTTCGATCAGAGTCATGCACCTG 13340
QY 200 ProProThrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGlu 219
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 13341 ---GCCCTGGCGGCGCATCTCTGTCGACGAGGTGACAGTTCGCGCTCGCGCATGCTCGAC 13397
QY 220 PheLeuAsnSerTyrHisGlnGlnAlaAspTyrLeuSerThrHisTyr-----AlaGlu 237
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 13398 TTGCTGGCG---ATCCATCTTCGAAGGATCGTGCACCAACCATTCGATTCGATTCGATTCAG 13454
QY 238 GlnGlyAsnHisArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyValSerPro 257
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 13455 GACAAACATACCGGTACCTCCGAGCGCCGACG---TGTTCATTGTTGGCAGTTGGCTT 13511
QY 258 GluPheLysAspSerPro-----ArgTrrpArgGlnThrGlyLeuSerValLeuAsn 274
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 13512 CGGGCTACGGCGAGCGCTTACGCCGCTTACCTGGATGCGGTAGGGAGAGAAATGGCTGGAA 13571
```

```
QY 275 ThrGluIleLysLysGlnValTyrAlaAspGlyMetClnPheGluLeuSerProIleTyr 284
Db 13572 AACCGGGCGAATCATCTGATTGGCGAGGAGCGGAGTTCAGCAGATTCGCTGAATTAC 13631
QY 295 HisValAlaAlaIleAspIlePheLeuLysAlaTyrGlySerAlaLysArgValAsnLeu 314
Db 13632 CACCGGGTCATGCTCGATACGTTCTCTACTGCGCAGAGTGGAGGCGTCAGCTTCCCTT 13691
QY 315 GluLysGluPheProGlnSerTyrValGlnThr-ValGluAsnMetIleMetAlaLeuI 334
Db 13692 GTGGAATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 13751
QY 334 eSerIleSerLeuProAspTyrAsnThrProMetPheGlyAspSer----- 349
Db 13752 TTCATTCAGCCCTGCCACT-GGTACGGTCCGAACCTTTGGGCTTAACGANGGTGCACGACT 13810
QY 350 ----TriIleThrAspLysAsnPheArg----- 359
Db 13811 GCTACCTCTGACCCAGACCGATTATCGTGACTATCGACCTGCGTGCAGTGGCCATGCG 13870
QY 359 aGlnPheAlaSerTyrAlaArgValPheProAlaAsnGln-----AlaIleLys 375
Db 13871 ATTGTTTCAACCGAT---ACCAACGCCCTATCCCGCAATCAAGAGTGCAGCCGCGTCT 13927
QY 375 sTyrPheAlaThrAspGlyLysGlnGlyLysAlaProAsnPheLeuSerLysAlaLeuSe 395
Db 13928 GTGGCTGGGGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 13987
QY 395 rAsnAlaGlyPheTyrThrPheArgSerGlyTyrAspLysAsnAlaThrValMetValle 415
Db 13988 CGAGGGTGGCTACGCCATTCTCGGAGCGGG-----CAGCAGATGCCAAT 14032
QY 415 uLysAlaSerProGlyGluPheHis---AlaGlnProAspAsnGlyThrPheGluLe 434
Db 14033 GCTGCGTTTACCCAGGTTCGTTTTCGCCCGCAGCCAGCGAGTCTTTCATCTTCATCT 14092
QY 434 uPheIleLysGlyArgAsnPheThrProAspAlaGlyValPheValTyrSerGlyAspGl 454
Db 14093 TTGGCTCGAGGGCGAAATCTGTCGCGGAGTCCGCGCAGTACAGTACACACCGCAATC 14152
QY 454 uAlaIleMetLysLeuArgAsnTyrTyrArgGlnThrArgIleHisSerThrLeuThrLe 474
Db 14153 CGCTTGGCTG-----AGTTACTTCCCGAGCGCCAGTCAACACAGATCCAGTT 14203
QY 474 uAspAsnGlnAsnMetVal-----lleThrLysAlaArgGlnAsnLysTyrGluThrGl 492
Db 14204 CGATGTCGCGATCAGATGCCCGCTCTCAGCCGTTTCTTTTGGCGACTGTTGAAGGC 14263
QY 492 yAsnAsnLeuAspValLeuThr----- 499
Db 14264 GACGAGGCTGTGCCCTGACTCGGAGCGAGCGTGGAGAAATGAATGATTTTCGGC 14323
QY 500 ----TyrThrAsnProSerTyrProAsnLeuAspHisGlnArgSerVal---LeuPheIl 517
Db 14324 GGCCTATACCGAT---GGGTAT---GGCGCAGGATAGCGGAGGTTCCCTTGTTCC 14377
QY 517 eAsnLysLysTyrPheLeuValIleAspArgAlaIleGlyGluAlaThr----- 533
Db 14378 GGATGGA-----CTCAGGTTATCGATGAGATTTCGGGCTTTTCGGCGAGTCCCGTCT 14431
QY 534 -----GlyAsnLeuGlyValHisTrp-----GlnLeuLysGl 544
Db 14432 GCGCTGGAGGCTGGACCTGGAG-CTGGTGTGTGGATGAAACACAGCTCAAGGGAGAGA 14490
QY 544 uAspSerAsnProValPheAspLysThrLysAsnArgValTyrThrTyrArgAspGl 564
Db 14491 GGGGAGAGCTGCTGATTTTCGAAGCTCGAT-----GCCCTCTTCGCTATCGACTGA 14544
QY 564 yAsnAsnLeuMetIleClnSerLeuAsnAlaAspArg----- 576
Db 14545 CAATCGGGTGGAGTCCGCTACTACTCTTCGGAAGAGAGGAGGTTCCTGTCTATCGATAG 14604
```

```
QY 577 -----ThrSerLeuAsnGluGluGly-----LysValSerTy 588
Db 14605 AGTTGGCGAGCGGGAACATTTGAACGGATTACAGTGGGCCCCGATGAGATCATCTA 14664
QY 588 rValTyrAsnLysGluLeuLysArgProAlaPheValPheGluLysProLysLysAsnAl 608
Db 14665 TTTCATCAACAC-----TTTTCAGTCCGAAGGATGGT 14700
QY 608 aGlyThrGlnAsnPhe 613
Db 14701 AGGCATCCGCTCCTAC 14716

RESULT 10
BD003866/c
LOCUS BD003866 3763 bp DNA linear PAT 31-JAN-2002
DEFINITION Polynucleotide of Streptococcus pneumoniae and sequence.
ACCESSION BD003866
VERSION BD003866.1 GI:18631827
KEYWORDS JP 2001501833-A/186.
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 3763)
AUTHORS Kunsch,C.A., Choi,G.H., Dillon,P.J., Rosen,C.A., Bara,S.C.,
Rannon,M. and Dougherty,B.A.
TITLE Polynucleotide of Streptococcus pneumoniae and sequence
JOURNAL Patent: JP 2001501833-A 186 13-FEB-2001;
HUMAN GENOME SCIENCES INC
COMMENT OS Unidentified
PN JP 2001501833-A/186
PD 13-FEB-2001
PF 30-OCT-1997 JP 1998520718
PR 31-OCT-1996 US 60/029960
PI CHARLES A KUNSCH, GIL H CHOI, PATRICK J DILLON, CRAIG A ROSEN, PI
STEVEN C BARASH,
PI MICHAEL FANNON, BRIAN A DOUGHERTY
PC C12N15/09,A01K67/027,C07K14/315,C07K16/12,C12N1/15,C12N1/19,
PC C12N1/21,
PC C12N5/10,C12P21/02,C12Q1/68,G06F17/30,C12N15/00,C12N5/00, PC
G06F15/40
CC Strandedness: Double;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1..3763
FT /organism="Unidentified".
FT /db_xref="taxon:32644"

BASE COUNT 1178 a 804 c 677 g 1104 t
ORIGIN

Alignment Scores:
Pred. No.: 7.21e-05 Length: 3763
Score: 176.00 Matches: 116
Percent Similarity: 34.63% Conservatives: 80
Best Local Similarity: 20.49% Mismatches: 231
Query Match: 5.04% Indels: 140
DB: Gaps: 20

US-09-802-285A-2 (1-659) x BD003866 (1-3763)

QY 71 SerLysAlaArgGluProAspPheSerAsnAlaGluLysProAlaAspIleArgGlnPro 90
Db 2173 TCAAAAGACTATCAGAAAGTCAAGAGTTCCTTGAACGCTTGATGCACATCGTTTATG 2114
QY 91 IleAspLysValThrArgGluMetAlaAspLysAlaLeuValHisGlnPheGlnProHis 110
Db 2113 TTTGATAGT---CCTGGGATGAGGCGCTGTTCAAAATCCATCAATCCAGCGG--- 2060
QY 111 LysGlyTyrGlyTyrPheAspTyrGlyLysAspIleAsnTrpGlnMetTrpProValLys 130
Db 2059 -----ATGGTATGGGATCAAGTATTTGAAGAT 2033
```

QY	131	AspAsnGluValArgTyrPheGlnLeuHisArgValLysTyrPheGlnAlaMetAlaLeuVal	150	QY	458	LysLeuArgAsnTyrPheGlnThrArgGlnLeuHisSerThrLeuThrLeuAspAsnGln	477
DB	2032	GATCCAGATGCTTTATATGCTCAATCGAAGAATATCTCTTCAGTTTATGATAGG	1973	DB	1037	TCTCTAAAGAGTCTTGGAGTCATTCGACCTGCATTGT	999
QY	151	TyrHisAlaThrGlyAspGluLysTyrAlaArg	163	QY	478	AsnMetValIleThrLysAlaArgGlnAsnLysTyrPheGlu	496
DB	1972	TATCTGGTAGAAGGAGATAAGGACTATATTCAAAGTCAAGTCTCTTCTATTTGATTGG	1913	DB	998	AGATGGGAAAGCTCCGGAAGAATCACGGG	969
QY	164	ValTyrGlnTyrSerAspTyrPheLysAlaArgLysAsnProLeuGlyLeuSerGlnAspAsn	183	QY	497	ValLeu-ThrTyr	500
DB	1912	ATTGACGAGGTGAGAGAATTTCTCTCAATCTTGATGACTAGAACCTTGATGATGGGT	1853	DB	968	ATCCTTGGGAATATGAATACTATCTCCTCTGTTTGTTCACATAAAGAAGGAGGG	909
QY	184	LysPheValTyrArgProLeuGluValSerAspArgValGlnSerLeuProPro	201	QY	501	ThrAsnProSerTyrProAsnLeuAspHi	510
DB	1852	ATTCGTTCTTTACTTGGTGAACATA	1826	DB	908	AGTGCATTATATTAGGGGGCTTATTTGGTCACAGAACCTGATTTCCTTATCTT	852
QY	202	ThrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGluPheLeu	221	QY	510	scnArgSerValLeuPheIleAsnLysLysTyrPheLeuValIleAsp	528
DB	1825	CTCTTGCTCTCTTGAATTTGACTTTGCTAGAGAGAGAAAGAACTAGAGAAATTTTG	1769	DB	851	CAGGAGAAAAATCCTCATGTTGGTAGAGGATCTCTGGCTCTTGTGTAGATGACATCAGGTG	792
QY	222	AsnSerTyrHisGlnGlnAlaAspTyrLeuSerThrHisTyrAlaGluGln	238	QY	528	atleGlyGluAlaThrGlyAsnLeuGlyValHisTyrPheGlnLeuLysGluAspSerAsnPr	548
DB	1768	GTCAGTCTAGAAAAGCAGATGACTTTATGAAAAGCTACTATCGCGCCCAAGTACACCCCT	1709	DB	791	TCAGGTCAA	761
QY	239	GlyAsnHisArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGlu	258	QY	548	oValPheAspLysThrLysAsnArgValTyrThrTyrArgAspGly	566
DB	1708	AGTAAGTGGGATTTTACAAATTCGATCGCTGTCTATCTATCATTTCTTTTCAGAT	1649	DB	760	ATCCTTGACAGGAT	717
QY	259	PheLysAspSerProArgTyrPheGlnThrGlyIleSerValLeuAsnThrGluLeuLys	278	QY	566	nLeuMetIleGlnSer	571
DB	1648	AAGATGGATCTAGAAAGCTTACCATTGCTTCAGAGAGTGAACACAGAAATTTGAG	1589	DB	716	GTTGAGACTATGGAGT	701
QY	279	LysGlnValTyrAlaAspGlyMetGlnPheGluLeuSerProIleTyrHisVal	296	RESULT 11			
DB	1588	ACACAGATTTTAGGAGATGGAAGCCAGTTTGAACAGTCAATCTCTATCATCTAGAGT	1529	AE007344	14385 bp	DNA linear	BCT 31-AUG-2001
QY	297	AlaAlaIleAspIlePheLeuLysAlaTyrGlySerAlaLysArgValAsnLeu	314	LOCUS			
DB	1528	TATAAAGCTTCTGATTTGTCTC	1499	DEFINITION			
QY	315	GluLysGluPheProGlnSerTyrValGlnThrValGluAsnMetIleMetAlaLeuIle	334	VERSION			
DB	1498	CTTCCAGACTTGCAGATAGTTACCAAGAGTTGCTGGAAGAGTGGCGACCTATATCAA	1439	KEYWORDS			
QY	335	SerIleSerLeuProAspTyrAsnThrProMetPheGlyAspSerTyrPheThrAspLys	354	SOURCE			
DB	1438	ATGATGACAGGCTTAGATGGACGGACTTTGGCTTTTGGTATAGCGATTCTACAGAAACG	1379	ORGANISM			
QY	355	AsnPheArgMetAlaGlnPheAla	362	REFERENCE			
DB	1378	ACAGAAATTTTGACCTGCTGCTGTTGTTGACAGGAGACCTTCTTAACGGTCTG	1319	AUTHORS			
QY	363	SerTrpAlaArgValPheProAlaAsnGlnAlaIleLysTyrPhe	377				
DB	1318	GATGTTAAAGTTGATTTGCTAGCTCTTGTCTCTGGGGGAGAAAGCTCAAGCGACTG	1259	TITLE			
QY	378	AlaThrAspGlyLysGlnGlyLysAlaProAsnPheLeuSerLysAlaLeuSerAsnAla	397	Complete genome sequence of a virulent isolate of Streptococcus			
DB	1258	CAGGAATTTGAAGAGCTTGGCGACCTAG-----TCCATGATCTTTGAAGACTCT	1205	pneumoniae			
QY	398	GlyPheTyrThrPheArgSerGlyTyrAspLysAsnAlaThrValMetValLeuLysAla	417	JOURNAL			
DB	1204	GGACATGCTGCATTAAGGAT-----GAACATCGTTATCTATTTTCAAAAT	1157	Science 293 (5529), 498-506 (2001)			
QY	418	SerProProGlyGluPheHisAlaGlnProAspAsnGlyThrPheGluLeuPheLys	437	MEDLINE			
DB	1156	GTCTCGCTAGGAAGTCCCATGATCATAGTACGAGAAATAGTTTTTGTCTACATATCAA	1097	21357209			
QY	438	GlyArgAsnPheThrProAspAlaGlyValPheValTyrSerGlyAspGluAlaIleMet	457	PUBMED			
DB	1096	GGCCACCTATTTTCATAGATGCTGGCGGTTATCTTATCG-GGAGATATATGAACGTTA	1038	11463916			
				REFERENCE			
				AUTHORS			
				TITLE			
				JOURNAL			
				FEATURES			
				source			

```

/organism="Streptococcus pneumoniae TIGR4"
/db_xref="TIGR4"
/db_xref="taxon:170187"
/complement(142..618)
/gene="SP0313"
/complement(142..618)
/gene="SP0313"
/notes="identified by match to PFAM protein family HMM
PF00255"
/codon_start=1
/transl_table=11
/product="glutathione peroxidase"
/protein_id="AAK74490.1"
/db_xref="GI:14971787"
/translation="WTSLLYDFSVLNNQNAOTPLDSYRGKVLIVNTATCGTLPQVQV
LDELRYQDQFEILDPCFMGQAPGSAEINAFCSLHFQTFPFRFKIKVNGKE
ADPLVYMDQSGPLGRVFNFAKFLIGRDGVFERFSSTDPKQIEAIIQTLL"
841..4041
/gene="SP0314"
/notes="identified by match to PFAM protein family HMM
PF02278"
/gene="SP0314"
/notes="identified by match to PFAM protein family HMM
PF02278"
/codon_start=1
/transl_table=11
/product="hyaluronidase"
/protein_id="AAK74491.1"
/db_xref="GI:14971788"
/translation="MQTKTKLVSLSSVLSSGLNLNHYMTGABETTNTTIQSQOKE
VYQORDTKNLVNGSDGDSGSPWTGSKAQGSAAWVDQKNSADASTRVIEAKDGA
ITTSHEKRAALAHMRVPTAEAKKYLRFKIKTDNKGIAKVRITIEESQKDLKNSA
TTSQTKDMQTIADYSPTLDVDKIKLEFVETGCTVSEKDIELVEADOLSEDSQTD
KQLEBKIDLPQKXVFLADYTYKVNFEDVASVNGILPELKEGTTNNIVSKDQREV
KXIPKILASVDAYTDRLDWNGLIAGQYVDSRNEQAKNLQELGKVAQSLSSIS
SQADRTYWEKSNYKTSANLTATYRKLEMAKQVTPSSRYQDETIVRTYRDSMEW
MEKHVYNQKSTVGNWMDYEIGTPRAINNTLSIMKEYFDEEIKYTDVIEKFPDPE
HEPRTDNPFKALGNLDMGRVGIAGLRKDDOEISSTIRSEOVFKLVQGGFY
QDGSVIDHTNAVYTAGYNVLIDGLSOLLPVQTKNPDKDKMOTMWHIDKSPAL
LVNGLMDMSGRSISRANSEGHVAEVLRIHRIADMSGETQKQCSLVKTIQVS
DSYDVKTKYTKDLSMQSLSDAGVASVPSYLSAFNMKDMTAMNAEKGFGF
LSLFSRTNLNHYMKENKRGWYTDGMPYLYNGDLSHYDGYWKTNPYKMPGTET
DAKRAEDSTGTPSALFVGTSKLDDANATMDFTNWNQTLTAHKSWMFKDKIAPLG
SNIQNTSDTAATTIDQRLSEGNPKVYVNDKEASLTQEKDYPTQSVLESDFSK
KNIGYFFPKKSI SMSKALQKAWKDINEGQSKSENEFELTISOAHKQNRSDGYML
INVDRAFPQMKELLESLLIENNETLQSVYDAKQGWGIVKYKIDSVESTISQFQVLK
RGVYTRKGGDEYKATYNPETOESAPDOEVFKKLEQAAQPOVQNSKEKEKEBEKNH
SQQKMLPQKGQCSILASGLGLLGAFLYLRGRKN"
4425..5045
/gene="SP0315"
/notes="similar to SP:Pl9780 GB:Y00434; identified by
sequence similarity; putative"
/complement(5058..5165)
/gene="SP0316"
/complement(5058..5165)
/gene="SP0316"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK74492.1"
/db_xref="GI:14971789"
/translation="WVYLRASIPNHQPAKDGAFHVQVQALLLTLRHLPL"
/complement(5189..5818)
/gene="SP0317"
/complement(5189..5818)
/gene="SP0317"
/notes="identified by match to TIGR protein family HMM
TIGR01182"
/codon_start=1
/transl_table=11
/product="4-hydroxy-2-oxoglutarate
aldolase/2-dehydro-3-deoxyphosphogluconate aldolase"
/protein_id="AAK74493.1"

```

```

/db_xref="GI:14971790"
/translation="MTKSDTIIEELKXQKIVAVIRGNTKEEGLQASTACIKGGIKAEI
AYNQIAGQIIKELVDLQDDQSCVICGAGTVLDVATARDAILAGNAVVPSPFHAETA
RMCNLYSTPIYIPGCTITLITALEAGSEIILKFPSTLSPAYISAVKAPIQVSVMV
TGGVGLNNIPQWFAAGADAVGIGELNKLASQGNFDRISEIAQQYITLR"
/complement(5828..6829)
/gene="SP0318"
/complement(5828..6829)
/gene="SP0318"
/codon_start=1
/transl_table=11
/product="carbohydrate kinase, pfkB family"
/protein_id="AAK74494.1"
/db_xref="GI:14971791"
/translation="WTKILLPGEPLIRISPLDATSIGDHVASTVPFGGSEINIACNLQ
ALGISTKPTALPANEIGDRFLTIKQHQIDTSSICRLGDRIGLYLENGFCGRQSEV
FYDRHTTISQIRPNMLDMDSLFQISHFSGITVAIGQEVRAILLILLLEAKRGGI
VYSMLNLRMTMISLEAKYFSPKFTDYCFGLDPLMDQNLMPFDRDASLEEV
ENRMILLKEAYGFKALFHTLRSSDRQDNVQYQALELFEESVOLKTAAYORIGSGD
AFISGALYQLLHSSKLTITLTPAVASATLKTCLPGDHLSTSTSTENLLANAQDIIR"
/complement(6860..7501)
/gene="SP0319"
/complement(6860..7501)
/gene="SP0319"
/codon_start=1
/transl_table=11
/product="conserved domain protein"
/protein_id="AAK74495.1"
/db_xref="GI:14971792"
/translation="MKIALINENSOAKSNHIIYDSLKEATDKKQVLFNYGMRGEGE
SOLTVQNGLMAAILNNTKAVDFVTGCTGVGAMALNSFPVCGGLAVDPTDVLV
SOINGNALISPYAKGFGWGAELTLKMFERLPAEMGGGYPRERVIPQRNARILNE
VQIITHNDLITLKIIDQFLDKTISGTFQYFFENCODDEVAAYLKEVLAK"
/complement(7520..8335)
/gene="SP0320"
/complement(7520..8335)
/gene="SP0320"
/codon_start=1
/transl_table=11
/product="oxidoreductase, short chain
dehydrogenase/reductase family"
/protein_id="AAK74496.1"
/db_xref="GI:14971793"
/translation="MTNITSFSTEQSLKGLKIALITGASVIGIGFAITAKSVAEAGATIVF
NDINQDLNKGIEAYREVGIVQAHGVCDVDEDDGQAMVKQIEQEVGVIIILVNNAGI
IRRVPMCSAADFRKVIDIDLNAPFIVSKAVIPSMIKKHGKIINICSMGBELGRET
VSAYAAAKGGLKMLTRNTIASEYGGANIQCNGIPGVIATPQTPAPLRELOQDSRHPFD
QPIIAKTPAARWGNTEDLMPAVFLASDASNFVNGHILYVDGGIILAYIKQPE"
8606..9040
/gene="SP0321"
/notes="similar to GP:5869507; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="PTS system, IIA component"
/protein_id="AAK74497.1"
/db_xref="GI:14971794"
/translation="MKVILVGHGFATGIYSSLSQLIAGNQENVEAIDFVEGMSADELK
QKILLAINEEVEVLLSDLGSPFKVSTINGENPAKTMVNLGSLNLAAMEAVFAR
NAHSFDEVVNVKSVAAQGGVNGKELFSTDAEEEDFESGI"
9052..10242
/gene="SP0322"
/notes="similar to GP:5869507; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="glucuronyl hydrolase"
/protein_id="AAK74498.1"
/db_xref="GI:14971795"
/translation="MIKVTIEKIKSPERFLEVPILLTKVEVGQAIDKVIROJELNLDY
FKEDFTPATFDNVYPIMDNTIEWNGFTWELWLALEYSSQDAPFKNIAHKNVLSFLDR

```

VNKRVELDHDHDLGFLYTPSCMAEYKINGDEAREATLKAADKLIERIOYKGGFIOAWG
DLGKPEHRLILLIDCLNLQILFFAEOETGDKYDYIAESHEFYASANNVIRDDASSFHT
FYEPDQPGKPTQROYSDDSCWARGOSNGVIGIPLITYRHLDKDESCFDLFKGVINY
FINLPKDHVSYDWLIFNDGSDQSRDSATAIAVCGIHEMLKHLUPEDVADKDITYKHAM
HMLRSLIHEVANDQFTPGGTSLLHGVYSHWSKGVDEGNIWGDYIYLEALIRFYKOW
NLYW"

gene 10253..10744

/gene="SP0323"

10253..10744

/gene="SP0323"

/codon_start=1

/transl_table=11

/product="PTS system, IIB component"

/protein_id="AAK74499.1"

/db_xref="GI:14971796"

Alignment Scores:

Pred. No.:	0.000429	Length:	14385
Score:	176.00	Matches:	116
Percent Similarity:	34.63%	Conservative:	80
Best Local Similarity:	20.49%	Mismatches:	231
Query Match:	5.04%	Indels:	140
DB:	1	Gaps:	20

US-09-802-285A-2 (1-659) x AB007344 (1-14385)

Qy	71	SerLysAlaArgGluProAspPheSerAsnAlaGluLysProAlaAspIleArgGlnPro	90
Db	12733	TCAAAGACTATGAGAAGTCAAGAGTCCCTTGGAACGCTTGATGGACAATCGTTTTATG	12792
Qy	91	IleAspLysValThrArgGluMetAlaAspLysAlaLeuValHisGlnPheGlnProHis	110
Db	12793	TTTGATAGT---CCTTGGGATATGGAGCTTGTTCAAATAATCCAAATCCAGCGC---	12846
Qy	111	LysGlyTyrGlyTyrPheAspTyrGlyLysAspIleAsnTrpGlnMetTrpProValLys	130
Db	12847	-----ATGGTATGGGATCAAGTATTTCAAGAT	12873
Qy	131	AspAsnGluValArgTrpGlnLeuHisArgValLysTrpTrpGlnAlaMetAlaLeuVal	150
Db	12874	GATCAGAAATGGTCTTATATGCTCAATCGAAGAATATCTTTCAGTTTATGATAGGG	12933
Qy	151	TyrHisAlaThrGlyAspGluLysTyrAlaArg-----GluTrp	163
Db	12934	TATCTGGTAGAAGGAGATAAGGACTATATTCAAAGTCAAGTCTTCTTCTATTTGATGG	12993
Qy	164	ValTyrGlnTyrSerAspTrpAlaArgLysAsnProLeuGlyLeuSerGlnAspAsnAsp	183
Db	12994	ATTGACGAGGTGAGAGAATTTCTCTCAATCTTGATGACTAGAACCTTGGATACGGGT	13053
Qy	184	-----LysPheValTrpArgProLeuGluValSerAspArgValGlnSerLeuProPro	201
Db	13054	ATTCGTTCTTTACTTGGTTGNAACTA-----	13080
Qy	202	ThrPheSerLeuPheValAsnSerProAlaPheThrProAlaPheLeuMetGluPheLeu	221
Db	13081	---CTCTTGCCTCTTCAAATTTGACTTGCTAGAGGAGAAAGAACTAGAGAAAATTTG	13137
Qy	222	AsnSerTyrHisGlnAlaAspTyrLeuSerThrHisTyrAlaGluGln-----	238
Db	13138	GTCAGTCTAGAAAAGCAATGACTTTTATGAAGACTACTACGCCCAAGTACACCCCT	13197
Qy	239	GlyAsnHisArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGlu	258
Db	13198	AGTAACCTGGGATTTTACAACAATTCGATGCTTGCTATCTATCTATCTTTTCAGAT	13257
Qy	259	PheLysAspSerProArgTrpArgGlnThrGlyIleSerValLeuAsnThrGluLys	278
Db	13258	AAGATGGATCTAGAGAAGCTTACCATTTTGCTTCAGAGGAGTTGAAAACAGCAAAATTCAG	13317
Qy	279	LysGlnValTyrAlaAspGlyMetGlnPheGluLeuSerProIleTyrHisVal-----	296
Db	13318	ACACAGATTTTAGGAGATGAAGCCAGCTTTGAACAGTCAATCTCTATCATGTAGAGGTT	13377

Qy	297	-----AlaAlaIleAspIlePheLeuLysAlaTyrGlySerAlaLysArgValAsnLeu	314
Db	13378	TATAAAGCTTCCTGGATTTTGTCTC-----TTG	13407
Qy	315	GluLysGluPheProGlnSerTyrValGlnThrValGluAsnMetIleMetAlaLeuIle	334
Db	13408	CTTCAGACTTGCAGATAGTTTACCAAGAGTGTCTGGAAGATGCGCACCTATATCAA	13467
Qy	335	SerIleSerLeuProAspTyrAsnThrProMetPheGlyAspSerTrpIleThrAspLys	354
Db	13468	ATGATGACAGGCTTAGATGAGCGGACTTTGGCTTTTGGTGATAGCATTCTACAGAAACG	13527
Qy	355	AsnPheArgMetAlaGlnPheAla-----	362
Db	13528	ACAGAAATTTTGAGCCTCTGCTGTGTTTGAACAGGAGACCTCTTAAACGCTCTG	13587
Qy	363	-----SerTrpAlaArgValPheProAlaAsnGlnAlaIleLysTyrPhe	377
Db	13588	GATGTTAAAGTTGATTTGCTTAGCCTCTTGTCTGGGGCGAGAAAGGTCAAGCGACTG	13647
Qy	378	AlaThrAspGlyLysGlnGlyLysAlaProAsnPheLeuSerLysAlaLeuSerAsnAla	397
Db	13648	CAGAAATTTGAAAAGAGAGCTTGGCAGCCTAAG-----TCCATGATCTTTGAAGACTCT	13701
Qy	398	GlyPheTyrThrPheArgSerGlyTrpAspLysAsnAlaThrValMetValLeuLysAla	417
Db	13702	GGACATGCTCGATTAGGAT-----GAACATCGTTATCTATTTTCAANAAT	13749
Qy	418	SerProProGlyGluPheHisAlaGlnProAspAsnGlyThrPheGluLeuPheLys	437
Db	13750	GGTCCGCTAGGAAGTCCCATAGTACAGCAGATACTTTTGTCTTACAGTATCAA	13809
Qy	438	GlyArgAsnPheThrProAspAlaGlyValPheValTyrSerGlyAspGluAlaIleMet	457
Db	13810	GGCCAAACCTATTTTCATAGATGCTGGCGGTTATTCTTATCG--GGAGATATATGAACGTTA	13868
Qy	458	LysLeuArgAsnTrpTyrArgGlnThrArgIleHisSerThrLeuThrLeuAspAsnGln	477
Db	13869	TCTCTTAAAGAGTCTTGGAGTCATTCGACCTGCATTGT-----	13907
Qy	478	AsnMetValIleThrLysAlaArgGlnAsnLysTrpGlu---ThrGlyAsnAsnLeuAsp	496
Db	13908	-----AGATGGGAAAGCTCCGGAAAGAAATCACCGG	13937
Qy	497	ValLeu-ThrTyr-----	500
Db	13938	ATCCTGGGAATATGAATACTATCCTCACTCCTCTGTTTGTCAACATAAAGAAAGGAGGG	13997
Qy	501	-----ThrAsnProSerTyrProAsnLeuAspHi	510
Db	13998	AGTGCATTATATTGAGGGGCTTATTGTCAGCAGAACCTGATTTGCTTATCTTT---CA	14054
Qy	510	sGlnArgSerValLeuPheIleAsnLysLysTyrPheLeuValIleAsp-----ArgAl	528
Db	14055	CAGGAAAAAATCCTCATGTTGGTAGAGGATGTCTGGCTTGTGGTAGATGACATCAGGTG	14114
Qy	528	alleGlyGluAlaThrGlyAsnLeuGlyValHisTrpGlnLeuLysGluAspSerAsnPr	548
Db	14115	TCAAGGTCAA-----CATGAGGTGTGACTCAGTTT-----	14145
Qy	548	oValPheAspLysThrLysAsnArgValTyrThrTyrArgAspGly-----AsnAs	566
Db	14146	-ATCCTTCACACAGGAT-----GTGACCTTCAAGATGGGAAAATCAATCA	14189
Qy	566	nLeuMetIleGlnSer	571
Db	14190	GTTGAGACTATGGAGT	14205

RESULT 12

AF498414

LOCUS

DEFINITION Pseudomonas aeruginosa serotype O3 putative O-antigen biosynthesis

AF498414 19295 bp DNA linear BCT 13-JUN-2002

gene cluster, partial sequence.
AF498414 AC104733
AF498414.1 GI:20560007
Pseudomonas aeruginosa.
Pseudomonas aeruginosa.
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
1 (bases 1 to 19295)
Raymond, C.K., Sims, E.H., Kas, A., Spencer, D.H., Kutayavin, T.V.,
Ivey, R.G., Zhou, Y., Kaul, R., Clendenning, J.B. and Olson, M.V.
Genetic variation at the O-antigen biosynthetic locus in
Pseudomonas aeruginosa
J. Bacteriol. 184 (13), 3614-3622 (2002)
22053227
12057956
2 (bases 1 to 19295)
Raymond, C.K., Sims, E.H., Kas, A., Spencer, D.H., Kutayavin, T.V.,
Ivey, R.G., Zhou, Y., Kaul, R., Clendenning, J.B. and Olson, M.V.
Direct Submission
Submitted (04-APR-2002) Genome Center, University of Washington,
Box 352145, Seattle, WA 98105-2145, USA
On May 13, 2002 this sequence version replaced gi:17975262.
Location/Qualifiers
1..19295
/organism="Pseudomonas aeruginosa"
/serotype="03"
/db_xref="taxon:287"
<1..104
/note="ORF 1; C-terminal coding region of *tpsa*"
/codon_start=3
/transl_table=11
/product="RpsA"
/protein_id="AAAM27759.1"
/db_xref="GI:20560008"
/translataion="EKDAMKELRQEVESAGPTTIGDLIRAQENQG"
241..525
/note="ORF 2; himD/hmfB"
/codon_start=1
/transl_table=11
/product="himD"
/protein_id="AAAM27760.1"
/db_xref="GI:20560009"
/translataion="MTKSELIERIVTHQGQLSAKDVELAIKTMLEQMSQALATGDRIE
IRGGSFSLHYRAPRVGNPNKTGSVRLDGKVFVHPKPKELRDVNEP"
559..861
/note="ORF 3; hypothetical orfa"
/codon_start=1
/transl_table=11
/protein_id="AAAM27761.1"
/db_xref="GI:20560010"
/translataion="MQWFRRAAVALIVSVIVVFLNQGVVQLAFVGVOFPWP
LAVYIIIAFILGILGLAQLPFLTISRVRASSRAELVQARKEVDTLRAEPSKVS"
1146..2174
/note="ORF 4; wzg; similar to chain length determinant
protein"
/codon_start=1
/transl_table=11
/product="Wzg"
/protein_id="AAAM27762.1"
/db_xref="GI:20560011"
/translataion="MSKNEIDFIELIQLWAQKWLVLSTLFGAVLAGSYAFSLKPV
YEARVAVLPSSLDVAGNIARSODGLPPTVSQVSVFTRNIOSEGTRRQRFREVY
LPSLERRVGSVDLGRFSDVNVNPTKELPERYTVFDRHDPDQAKWAKOYIELA
AORSMDERGNARQLELYKSNIQQOITLQVVKERRTRDRIYVRLKALQVARAGVE
KPEPLGSDQDLQSAIMSGMSLMRGSRALRAIEALEARTSDPFPALURGJOEQA
LYESKQLGDFRVAFRDQGSVETPDPIKPKKVLITLALGILAGVLGVALAILLFLFR
EXRLDITG"
2421..3491
/note="ORF 5; similar to UDP-N-acetylglucosamine
2-epimerase"
/codon_start=1
/transl_table=11
/protein_id="AAAM27768.1"
/db_xref="GI:20560017"
/translataion="MWKAGRYIIVGGFPDPVGVGTAFVSRLEANNWSEVIDIYPSI
P"
/protein_id="AAAM27763.1"
/db_xref="GI:20560012"
/translataion="MAPVTALAGDRFARVCTVGOHREMLDQVLFKTIKVPDFDLN
IMRTGQDITDITVALIQGMKEVFAQKPDIIILVHGDATTTLAAISLAAYVHOIVAHVE
AGLITGNMYSWPBEGNEKLTGALACILHAFETETSKONLAREGIAAARQIVVGTNTVIL
ALLSVQRLDHVDVLARQAAPTEFLDSSRFVLTGHRRESFGGFERICRAIAEVA
RNHPEDIVPVHLNPNVREPVRNLLSIVENHILIEIDYLPFIHLMRAHIILTDGSG
GIOEAPSLIGKPVLMVRETTTERPEAAVAGTVKLVGTEVKNIVNLELRLTDESAYLQ
SFAHPYGDGGEACQRIQALLV"
3523..4785
/note="ORF 6; similar to UDP-glucose/GDP-mannose
dehydrogenase"
/codon_start=1
/transl_table=11
/protein_id="AAAM27764.1"
/db_xref="GI:20560013"
/translataion="MSFQTIISVIGLIGVIGLTAAPVAFSARQKKVIGVDVNOAAVDITNR
GEIHIVEPDLDMVHAATEGLRATTNPEDAPFLIAVPTPFKHEHEPDLTLEAAS
KSIAPVLKKGDLVILESTSPVQATEQWAGLAEFRPDLTPFQTHGEASDRLVRAHCEP
VLPHGLRELVDNRVIGMSEKGESEAAKLYKIFKVEGCIITNARAEKMTCTNSF
RDVNFARLSARSIIICDKLDINWELIRLANRHPRVNLOPGCVGSHVADVDFWIVS
KTPDEARILARSERNDHKEPFWLEKILAVADVLOKPEKTARDVDSVACFGLAFAV
IDDLRESAIIAIAQVVASHPGRVIVVEPNIEELPGKSGNLTASTDSAMQEADVVV
LLVDHKEFEIEAPHRQLVVDITRLGMEA"
5040..6155
/note="ORF 7; similar to Polysaccharide biosynthesis
protein; potential multiple membrane spanning domains."
/codon_start=1
/transl_table=11
/protein_id="AAAM27765.1"
/db_xref="GI:20560014"
/translataion="MLIHALVGYASIFDGLGKSVWRQIAISAEPEKRGPIILGSSILL
FVLLSGGLAAALVHFFAFLVVDIFKVDSSYSDALGTGRAALCIPFFLASIVLQGY
LEGVEDFYSFNKRYALSGTLTLLPVAFLLEKFTFSLIYGLLPARAGSCFELALIVS
RCSVWTRWDVVKPKTLISFGSWLITITNIIISPVWYMDKFLARITAGELAYAAAP
SELINMSILPVAITRAVFPRLSALGGQSGASVKRQALLYSGGLVLDPIIILFFVAPW
GLIGMLPEYSGTAAQVQFQIMLVGLDFNSIALPYNLQAKYSRTTAMVHLLEVLFP
LMLMYFSVQGVGVGAASINSLRMAVDCILMAYLDSK"
6162..7412
/note="ORF 8; potential multiple membrane spanning
domains."
/codon_start=1
/transl_table=11
/protein_id="AAAM27766.1"
/db_xref="GI:20560015"
/translataion="MASYITEKSLKFLFFVLLTLMLSFPKSEFAVIKIVILAFMFSMVF
LHVLGVYRRFSLDIIIFLYFILLAVILWILVGLNDGYALIGDAPFLVVFSLFL
VYILWLENFDXLPILFISIMVSTIIIFAFCLSFIEALYSVLSFPPDLFVRLVNMAGV
HEGHQGLNNVSNLSLFTFPELVSALYSELKRYRFLFVFLVGLVIAVLLAGARRA
LMLILMVPLAFVFRIRISGVSSIEGGRGVKIIIFVTFPAVCGVGLIAGLVNNAQ
AERTESAFGDSVRVDQFISLWKGTDFPLLGSGFGGSVDVIRSEPRMVELSYMQL
LFNGSLGLFSLITGCIIGFYKALFVKIKNLSLYRAEAISITLVGLVFCVGSATNPYES
GDFSLLFLGFIPLIASVGKVAHS"
7399..8463
/note="ORF 9; similar to Glycosyl transferase"
/codon_start=1
/transl_table=11
/protein_id="AAAM27767.1"
/db_xref="GI:20560016"
/translataion="MRIREASMLKSNRVDPIVLVLLNMGWQDTIETCLESIRLHEQPO
DILVFDNGSENDYSVQLEERTISIAQHEEAFELDGHVLHI VHYDVSGVYRLCSIG
YFDRPEILWNCGVSVRWGRVYVYFSGKNTYLEIKEAPVGPATCCMLFKVDFFERS
IGGTEQPFPEEEDVELSPRLRRGTHVCAARAIYHKVGASIRGKDELI FRKTFI
YINRFINRRTOLPVAVVWNAWRFLVFKVAILTRNRNFKSGSSEILAFIRLLIMROSSRL
GVDKALFKVSLKSGLYNDG"
8456..9469
/note="ORF 10; similar to Glycosyl transferases group 1"
/codon_start=1
/transl_table=11
/protein_id="AAAM27768.1"
/db_xref="GI:20560017"
/translataion="MWKAGRYIIVGGFPDPVGVGTAFVSRLEANNWSEVIDIYPSI
P"

SKVPSBEKGRVPRKPTGFLSYLVIANHROWAGRLHFNSTSAIRVFFLVLPKFAS
RESLMHSGSLQVSKSRKMLYEAALSLRDLYVFCNKGHRQVFGILGVPEEKWRLVSSY
IKIPLSLISIVBERKDIQEFDFGGSVULVAGSPSSLYNHEWCLIDYYSAREGKLALFVLY
GSGDALNRIBERASERVKIIFWNCQATFNHALSRASVLRKTSKDSFGVAVADAVEL
GVVVASVCKRYPCAGLPVDFDFSAFCFTVDCILNGTDIIRTEDESGFRATFSYSLDL
RDV"

CDS

9466..11607
/note="ORF 11; similar to Zinc-binding dehydrogenases"
/codon_start=1
/transl_table=11
/protein_id="AAAM27769.1"
/db_xref="GI:20560018"
/translation="MKQVLQDMAGKGTILVTRAPAPQASKEHIVINSVLSLSAGTERM
LVGPGKASVLEKARQPEKVKMLVKQVTDGLMTTIEAVRSKLGLOPLGVCNVCVVS
EVGSGVEGKGVDRVNSGPHADVVKVSKNLCARIPDEVDDESAFVVFVAGTIGQIR
LAQPTLGEAFVGVVGLGLLTVQLLRHAGRCVLAIDYDENKLSIARGVGAATNPGFR
DEYPTAMATSRGGGVGVITASTQSDPVTAAHMRKRGRIVLVGVVTGLELNRRA
DFYKELTFQVSCSPGPRVDYNEEKQDYPVLFGRVTEQRFNFAILMLACGLDVP
QALITHREAFEDAPAOYALVDDKEAGLILLOVYTAASRAIRQLERQGVTFDPV
RVLGFIAGNVASMLIPAKAQAQFHTLATAGGIVNGHKGAFEAESTVDAM
LANPTINTVAVVTHDSHARFVSQVLRAGKSVFEKPLAIHAGELQEVRYDAYQAANSN
GASALQVGNFRFSPQVKMSLLAAVKEPKSFMTWNAGAIPIANHWITQKAVGGGR
IIGEACHFDLRLIAGSRIVSVQARSMDGAGVETEDKASITILFDGDFSGFTILYL
ANGAASFPERVEVETAGRVQLDNFRFKGYGWPFGSKLNLWKQDKGNACAKAFLD
GLQSQQAIPABEIEFVADVTIQVAELLRNQ"

CDS

11604..13448
/note="ORF 12"
/codon_start=1
/transl_table=11
/protein_id="AAAM27770.1"
/db_xref="GI:20560019"
/translation="MSSFMVKARTAWALGAPNLLVLAIRMVGMTGLNPVRRLOBRAP
SGDFPLTPPSPSVSLDKDQWLGKVFYVYADSDVDMHSLNLTGVGMTNSDRNW
WEIPDFDLAVGDIKAVWEASRPDVPALAQKALGESALDKLNALWLDWCMNPPYK
GNMKQGEASLRVWMLALAILQQVTAAPPALDLIRIHLRIVPTIRYALAQDN
HGTSAALFTGGSLVRHGEDPAVTWMLGRKWLLENRAKHLIGEDGFSQVSLNYHR
VMLDTFSLAEIWRQLSLVRSARWYSAATHWLHAFIDPATGDPNGLANDGARL
LPLTQTDVRYPAVOLAMALFTDNTNAYGQBCSRPLLWLGLDLPASBSVPVRSRVF
DEGGVAILRSGQAMAMLVPRPRPSPQADALHLDLWLEGNRLADACTVSYNTRSAF
LSYFPGTASHNTIQDGRDQMLSRFLFGDLWKATQAVPLTSSERASEMNAFSAAT
DQYGHAKREVALFPDGLKVIDEISGFSASVLRWLRPGSVLWLDGNQDILKRGDVLII
FESSMLLRILITIGESRYLAKEEVPIEVLEGQAGTLKTDYEWAR"

CDS

13445..14662
/note="ORF 13; similar to Glycosyl transferases group 1"
/codon_start=1
/transl_table=11
/protein_id="AAAM27771.1"
/db_xref="GI:20560020"
/translation="MRVITYFHQHFSTPKGMVGIIRSYEMARLVARGHQVWVWCGSYGG
GQTLDPVPSKRRQGTVDGIDIIEFDLAYSNTDGFIKRTTTFPSFALRSIRLAWTEE
YDLVATTPLTLAGIPGIFARWLKCAFVEVRDLWPELPKAMGVITNPVVLGMSVL
EMLSYSARHLIGLSPGIVEGIERLGVPAHLVEMVTPNGCDLIDIPASDSEPRPERVKA

Alignment Scores:

Pred. No.: 0.000634 Length: 19295
Score: 176.00 Matches: 137
Percent Similarity: 35.81% Conservative: 80
Best Local Similarity: 22.61% Mismatches: 244
Query Match: 5.04% Indels: 147
Dbs: 1 Gaps: 33

US-09-802-285A-2 (1-659) x AF498414 (1-19295)

QY 113 TyrGlyTyrPheAspTyr-----
Db 11817 TTCGGCTAGTCCCTATGCGCCTGATTCCTCGGATGGCACAGTAAATCTGCTGACT 11876
QY 119 -----GlyLysAspLeuAsnTrpGlnMetTrpProValLysAspAsnGlu 133
Db 11877 GGAGTGGGAATGACCAATTCTGATCGGAATTGG-----TGGGAGATCCCGGATTCGAT 11930
QY 134 ValArg-----TrpGlnLeuHisArgValLysTrpTrpGlnAla 146
Dbs: 1

Db 11931 CTGGCAGTGGGAGACATCAAGGCTCTCTGGGAGGCGTCCGCTTCGACTGG-----GTG 11984
QY 147 MetAlaLeuValTyrHisAlaThrGlyAspGluLysTyrAlaArgGlu----- 162
Db 11985 CTTGCTTTGGGCGCAAAAAGCTCTGGCTGGTGAGAGCTGGTGCGCTGCACAAACTGAATGCC 12044
QY 163 TrpValTyrGlnTyrSerAspTrpAlaArgLysAsnProLeuGlyLeuSerGlnAspAsn 182
Db 12045 TGGTTG-----AACGACTGGTGCAATGTAATATCTCTTACAAAGGGCCGGAAT--- 12092
QY 183 AspLysPheValTrpArg---ProLeuGluValSerAspArgValGlnSerLeuProPro 201
Db 12093 -----TGGAAATGGCGCCAGGAGCTTCGATCAGAGTATGACACCTGGCCCTG 12140
QY 202 ThrPheSerLeuPhe-----ValAsnSerProAlaPheThrProAlaPheLeuMetGlu 219
Db 12141 GCGGCCATTCTCTGTCAGCAGGTGACAGTCCG-----CCTGCCGCACTGCTCGAC 12191
QY 220 PheLeuAsnSerTyrHisGlnAlaAspTyrLeuSerThrHisTyr-----AlaGlu 237
Db 12192 TTGCTGGCG--ATCCATCTTCGAGGATCGTGCACCACTCGCTATCGGATGCTCAG 12248
QY 238 GlnGlyAsnHisArgLeuPheGluAlaGlnArgAsnLeuPheAlaGlyValSerPhePro 257
Db 12249 GACAACAATCACGGTACCTCCGAGGCGCAGCG---TTGTTTCATTGGTGGCAGTTGGCTT 12305
QY 258 GluPheLysAspSerPro-----ArgTrpArgGlnThrGlyLeuSerValLeuAsn 274
Db 12306 CGGGCTCAGCGGAGCCCTGACGCCCTTACCTGGATGCGGTAGGAGAGAAAATGGCTGGAA 12365
QY 275 ThrGluIleLysLysGlnValTyrAlaAspGlyMetGlnPheLeuSerProIleTyr 294
Db 12366 AACCGGGCGAAACATCTGATTGGCGAGGAGCGTACTTCAGCCAGTATTCGCTGAATAC 12425
QY 295 HisValAlaAlaIleAspIlePheLeuLysAlaTyrGlySerAlaLysArgValAsnLeu 314
Db 12426 CACCGGGTCATGCTCGATACGTTCTCACTTGGCAGCTGTGGAGCGCTGAGCTTCCCTT 12485
QY 315 GluLysGluPheProGlnSerTyrValGlnThr-ValGluAsnMetIleMetAlaLeuI 334
Db 12486 GTGGAAATCTCTGCTCGCTGGTATTTCGGCGGCAAGTCGAGCAGCTACCTGGCTGCACGCA 12545
QY 334 eSerIleSerLeuProAspTyrAsnThrProMetPheGlyAspSer----- 349
Db 12546 TTCATTGACCTGCCACT-GGTGACGGTCCGACCTTCGGGGTAAACGATGGTGCAGCT 12604
QY 350 ----TrpIleThrAspLysAsnPheArg-----MetAl 359
Db 12605 GCTACCTCGAGCCAGACCGGATTCGTCGACTATCGACTGCGGTGCGATTTGGCCATGGC 12664
QY 359 aGlnPheAlaSerTrpAlaArgValPheProAlaAsnGln-----AlaIleIly 375
Db 12665 ATGTTTACCGAT---ACCAAGCCTATCCCGGCAATCAAGAGTCGAGCCGACCGCTGCT 12721
QY 375 sTyrPheAlaThrAspGlyLysGlnGlyAlaProAsnPheLeuSerLysLeuAlaLeuSe 395
Db 12722 GTGGCTGGGGCTCGATCTGCTTCGCCAGTCTCTCTGTACCCGAGCGGAGTGTTCGA 12781
QY 395 rAsnAlaGlyPheTyrThrPheArgSerGlyTrpAspLysAsnAlaThrValMetVal 415
Db 12782 CGAGGGTGGCTACGCCATTCTCGCGAGCGGG-----CAGGCAATCGCAAT 12826
QY 415 uLysAlaSerProGlyGluPheHis---AlaGlnProAspAsnGlyThrPheGluLe 434
Db 12827 GCTGGCTTACCCAGGTTCGGTTTCGCCCGCAGCAGCGGATGCTCTTCATCTTGAATCT 12886
QY 434 uPheIleLysGlyArgAsnPheThrProAspAlaGlyValPheValTyrSerGlyAspG 454
Db 12887 TTGGCTCGAAGGGCGAAATCTGCTCGCGCATGCCGCGACCTACAGCTACACACCGAATC 12946
QY 454 uAlaIleMetLysLeuArgAsnTrpTyrArgGlnThrArgIleHisSerThrLeuThrLe 474
Db 12947 CGCTTGGCTG-----AGTTACTTCCCGGAACGGCCAGTCACACACGATCAGTT 12997

QY	474	uAspAsnGlnAsnMetVal	-----IleThrLysAlaArgGlnAsnLysTrpGluThrG1	492
QY	474	uAspAsnGlnAsnMetVal	-----IleThrLysAlaArgGlnAsnLysTrpGluThrG1	492
Db	12998	CGATGTCGCATCGATAGATCCCGCTCTCAGCGGTCTTTTGTGGCAGCTGTTGAAGGC	13057	
QY	492	yAsnAsnLeuAspValLeuThr	-----	499
Db	13058	GAGCGAGGCTGTGCCCTGACTTCGAGCGGACGCTCGAGGAATGATGATTTTCGGC	13117	
QY	500	----TyrThrAsnProSerTyrProAsnLeuAspHisGlnArgSerVal	----LeuPheI1	517
Db	13118	GGCCTATACCGAT--GGGTAT--GGCGCCAGGCATAAGCGGAGGTTCCTTGTTC	13171	
QY	517	eAsnLysLysTyrPheLeuValIleAspArgAlaIleGlyGluAlaThr	-----	533
Db	13172	GGATGGA-----CTCAAGGTTATCGATGAGATTTCCGGCTTTTCCGCGAGTCCGCT	13225	
QY	534	-----GlyAsnLeuGlyValHisTrp	-----GlnLeuLysG1	544
Db	13226	GGCTGAGCGTGGACCTGGAG-CTGGGTGTCGATGGAAACACGCTCAAGGAGAGA	13284	
QY	544	uAspSerAsnProValPheAspLysThrLysAsnArgValTyrThrTyrArgAspG1	564	
Db	13285	GGGAGACGTGCTGATTTTCGAAAGCTCGAT-----CCCTCTTCGGTATCGAGCTGA	13338	
QY	564	yAsnAsnLeuMetIleGlnSerLeuAsnAlaAspArg	-----	576
Db	13339	CAATCGGTGGAGTGGCGCTACTACCTTGGCAAGAGAGAGGTTCTCTCATCGAGATAG	13398	
QY	577	-----ThrSerLeuAsnGluGluGluGly	-----LysValSerTy	588
Db	13399	AGGTTGGGAGGGGGAACATTCAGACGATACAGTGGCGCCGATGAGAGTCATCTA	13458	
QY	588	rValTyrAsnLysGluLeuLysArgProAlaPheValPheGluLysProLysLysAsnAl	608	
Db	13459	TTTTCATCAACAC-----TTTTTCGACTCGAAGGAGGTGT	13494	
QY	608	aglyThrGlnAsnPhe	613	
Db	13495	AGGCATCGCTCCTAC	13510	
RESULT	13			
LOCUS	AF498401			
DEFINITION	AF498401	16720 bp	DNA linear	BCT 13-JUN-2002
ACCESSION	Pseudomonas aeruginosa serotype 010 putative O-antigen biosynthesis			
VERSION	AF498401	AC104720		
KEYWORDS	AF498401.1	GI:20559773		
SOURCE	Pseudomonas aeruginosa.			
ORGANISM	Pseudomonas aeruginosa			
REFERENCE	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;			
AUTHORS	1 (bases 1 to 16720)			
TITLE	Raymond,C.K., Sims,E.H., Kas,A., Spencer,D.H., Kutyavin,T.V., Ivey,R.G., Zhou,Y., Kaul,R., Clendenning,J.B. and Olson,M.V.			
JOURNAL	Genetic variation at the O-antigen biosynthetic locus in			
MEDLINE	Pseudomonas aeruginosa			
PUBMED	J. Bacteriol. 184 (13), 3614-3622 (2002)			
REFERENCE	22053227			
AUTHORS	1 (bases 1 to 16720)			
TITLE	Raymond,C.K., Sims,E.H., Kas,A., Spencer,D.H., Kutyavin,T.V., Ivey,R.G., Zhou,Y., Kaul,R., Clendenning,J.B. and Olson,M.V.			
JOURNAL	Direct Submission			
COMMENT	Submitted (04-APR-2002) Genome Center, University of Washington,			
FEATURES	Box 352145, Seattle, WA 98105-2145, USA			
source	On May 13, 2002 this sequence version replaced gi:17975249.			
	Location/Qualifiers			
	1..16720			
	/organism="Pseudomonas aeruginosa"			
	/serotype="010"			
	/db_xref="taxon:287"			

/transl_table=11
/protein_id="AA027559.1"
/db_xref="GI:20559780"
/translation="MWKLNNSUNALSLVAIQANALLPLLAPFLPGVLEQCAPAR
LVAAALAVYLVITVLSYDFDISGVQVIVDAKEEREVEAKFFNFPAARVGLLIIFSSVP
LVGGYVFWGADLVVILVFWLFGMLIQSNYFQAVENWNLATFVSSRTAVLAV
YFVHGKSTVFLASIIAGSLVGGAVGVLTRFGLSEBKIRGDDVVVLLMSGRH
LVFNGLVTLFRGANVILLAGVNSAAVAVLSEKIKTSQALARPUNOLFKPVVK
AWASISERRNAPAFRLVKNKTAVOLLVMEFLPVFLFLFYFCWVLGVIPGESEKVP
FLVLMSPAVIPFVNSMFGAVGLSLGGQYFAYSVSIIGVCAFLSLVMSYLLSDT
GAAMTVLAEALLLFAFLCKYGRPRHG"

CDS

6170..7504
/note="ORF_8; similar to u-PAR/ly-6 domain; potential
multiple membrane spanning domains."

/codon_start=1
/transl_table=11
/protein_id="AA027560.1"
/db_xref="GI:20559781"

/translation="MDEFFYLKLNIFLSSFFLLGAMCARKIFSVWLNPNVIFSLILWF
FYTVPLLVAFAPVSPFAYILYILFCVFLSLFVFPWGFAGRNENKPNVASCYFV
PLMVAFYSCSISAIMIFWGLQGLSLEOLLNPNVSIIGIYAGKYSGEIVSVVS
OVGLQCSYTAFFGLIYGARSEARGPKLLFFTEPALLWMLQSAKGLFFPSIFLIF
FGMLNVRNKNITLFGKDFRELLYGLLILPLIFSLRGLYQLDDTALMINRL
RYXIVTSSVHLPAFSDWFSERYFGESLMHYRQSSSTLGFYFMSFPLQAGDSDVAM
GVYDEFTYGEYVKNLYTVPRGIITDFGLLSGLFFAFILGLFNLCLVYRLCRNGSA
FAILFCYFVAISYQTYLISSLTWLTLPFVIVQWFMGLFLKFRFYTRSKV"

CDS

7501..8343
/note="ORF_9; similar to Glycosyl transferase"

/codon_start=1
/transl_table=11
/protein_id="AA027561.1"
/db_xref="GI:20559782"

/translation="WKICLPVCVNAHDDAVRELDVSNTAFSCGLGVELVILSDNST
LSSVITDARNYSNRYLQNSNIGYPPAPARALSTLTPEDISDFVWCVNDSLVSK
DFFIALSVSPCKDVIAPGIFSDKDRNPIMRPSLRKMKFMKMWVNSSTALFR
WYMLVARERTRACTORAKGGAIGQGPAMYGAGSFIITQLYPKRGAHVLYPR
FLGEEVFVAEQLNNMLIQHPDVRVSDKEHASTSKIKLFCPEHRKSYEYLLSK
FFDD"

CDS

8457..10625
/note="ORF_10; similar to Zinc-binding dehydrogenases"

/codon_start=1
/transl_table=11
/protein_id="AA027562.1"
/db_xref="GI:20559783"

/translation="WSCFIEGAMQVLQDMAKGGTSTVAPAPISRDHVLINVTSL
ISAGTERMLVGFGLYDLKARQOPEKVMVLEKIKTDGLMTVEAVQSKLAOLPLG
YCNVGVSEIGSGVEGPKVGRVNSGPHADLVKVKMNCARIIDGVDDESAFAVVA
SIGLQIRLQPTIGECFVTVGVLGILLTVQLRAHGRVLIADIDDDSKULELARQY
AQCNPQKGEDPVAAGMAFSRKGVDGVIITASTKSSDPVTOAARMKRKRRIVLGV
VGLNRADFEYKELTQVSCYGRDYDEYBKGGDPLGFRVWTEQRNFELVDM
LDSQLEVKTLITHRFPEDAPAYSTLTDQNSGLMLLYASDTAEVRVVRPLSDM
KGFNAQPVVGFICAGNYASRILIPAFKTAGAOPHTLVTSGINSVILHGEKAGFSQAS
TNIDAVNERMINIAIVTRHDTHARLVVDALRTGKNVFEKPLIAITHAELDVQAVH
AITHQGGPQLMGFNRFPAPHIQMKGLUASVREPKSFINTWAGAPVGHWTQDI
AVGGRIIGEACHPIDLMRYLGRIVSVQARRMGDSNAIDTSDEKASITLGFEDGSP
GTILYLANGASFPKRIEFTAGRVQLDNRFLKLGKYGWPGFNMKNMLWKDQKQROC
VAFLDSIEKGEPTIACEIEFEVTRVSEIAEAILRNQWD"

CDS

10634..12286
/note="ORF_11; similar to Prenyltransferase and squalene
oxidase"

/codon_start=1
/transl_table=11
/protein_id="AA027563.1"
/db_xref="GI:20559784"

/translation="WALQKLVLYLRYLLEKQIIFEVYVFAQVRYLYRVAVR
WERDPSRWNLNDVAVEFLGIGTKVQCQPDWQADRSKLLWNLHYLLDINAR
DIGQPQLQADLVQIQANPPVSGEGVEPYLSIRIVNLVKLARHDERSTFLIADSL
AVQADALQVVEYHILGNHLSFANGKALVFAGYLSGAMADRWLAKGLILDEELPEOF
LNDGHEFLSPMYHATLWDCNLSLRSGLEDAERLPQWREVYVQGLKWLRSWQ
HPDGRIISFNDAAFGIAPEYEDIAAYAKRLDISPAHENHLAALYNAGVAAVLPAD
GKAILDLAKVGPDPYOPGHADHTLSPFLSVFGKRLVNSGTSOYGDSDSRQORGTA
AHNTVGLLGYSDSSSEGVFRVRAAVVIEFEVEPILCTIQASHDGYGLVGLIRHR
RTWTFTRTSMEIFUSITGDQVATSRFPFSEVKVMKADGFVADMGKVKVSIIDFSG

SDDVRLISSTWHPFEGCSLANQCLVATFSGDSLTRICWGGI"
12355..13467
/note="ORF_12; similar to Glycosyl transferases group 1"
/codon_start=1
/transl_table=11
/protein_id="AA027564.1"
/db_xref="GI:20559785"
/translation="MSALGALPNSOLEVITTPNRYSSPLEAPLLEKHSRLIIHR
IPFOHNSGMLDQSRAPFSLKGVLTIRKNDVYVATSRSLTASIGALVABQLRC
PLYLDIRDFVDITIKDLPKRISFFVPIFSMMERIIVRTATRVNLSVAGFLPFNSR
YPGDFFVFTINGIDDEFVKVQTSFCESESSLYTLVYAGNMEGOGHLLIIPQIARHF
DHLRFLICIGGSRKQLODAIEAGCTNVEITSPVPSSELIAAYQADVLFELHNDY
DAFRKLPKIFGEYALKPFIWAGVAGYAGFIQDFISNAAVFPCNVEDAIFKAFSSL
SMVTKPRKFVERFAREKIMKALAEIDIFVVRGEVK"
13464..14417
/note="ORF_13; similar to NAD dependent
epimerase/dehydratase family"
/codon_start=1
/transl_table=11
/protein_id="AA027565.1"
/db_xref="GI:20559786"
/translation="MRVFLGITGFVGLATSKLLQEGHCVRAAVRDANPALAQVEQ
VFGNLVKNHWASSLEQGVVHCAARVHWKADSRPLSEFRVNVEGTINLARQAA
KAGVKRFVFISSIKVNGESTRLGPVYTADDIAMPELDFYGVSKLEAEQKLLSARVDM

Alignment Scores:

Pred. No.:	0.0015	Length:	16720.
Score:	170.50	Matches:	88
Percent Similarity:	36.61%	Conservative:	46
Best Local Similarity:	24.04%	Mismatches:	151
Query Match:	4.88%	Indels:	81
DB:	1	Gaps:	18

US-09-802-285A-2 (1-659) x AP498401 (1-16720)

Qy	144	TrpGlnAlaMetalLeuValTyrHisAlaThrGlyAspGluLysTyrAlaArgGluTrp	163
Db	10871	TGGCAAGCT-----GATAGGACGAGTAATAATTGG	10900
Qy	164	ValTyr-----GlnTyrSerAspTrpAlaArgLysAsnProLeuGlyLeuSerGlnAsp	181
Db	10901	CTCTATAATCTGCACATATTGGATGATCTGAACGCTAGGACATTGGC---AGTCAGCCT	10957
Qy	182	Asn-----AspLysPheVal-----	186
Db	10958	GGGTAGCCGATAAATACTAAATCAAGTTGGATCCAGGCTAATCCACCGTATCCGCTGAA	11017
Qy	187	---TrpArgProLeuGluValSerAspArgValGlnSerLeuProThrPheSerLeu	205
Db	11018	GGTTGGGAGCCTTATCTCTTCTTCTGCTATCGTAACTCTTGAAATGGCTGCTCGC	11077
Qy	206	PheValAsnSerProAlaPheThrProAlaPheLeuMetGluPheLeuAsnSerTyrHis	225
Db	11078	CATGATGAGCGTTTCACCTTT-----CTGGCGGACAGTCTAGCT	11116
Qy	226	GlnGlnAlaAspTyrLeu-----SerThrHisTyrAlaGluGlnGlyAsnHisArgLeu	243
Db	11117	GTACAAGCGGACGCGCTGGTGCAACAAGTAGAATATCATATCTTGGGAAATCATCTATTC	11176
Qy	244	PheGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPheLysAspSerPro	263
Db	11177	GCTAATGGGAAGCGTTGGTGTGGCGGC---GCTTATCTGTGAGCGCGCATGCTGAT	11233
Qy	264	ArgTrpArgGlnThrGlyIleSerValIleAsnThrGluIleLysGlnValTyrAla	283
Db	11234	CGTTGGCTTGCACAGGGGTACCGGATCTACTTCGACGAGAGCTCTCTGAGCAATTTCTGAAT	11293
Qy	284	AspGlyMetGlnPheGluLeuSerProIleTyrHisValAlaIleAspIlePheLeu	303
Db	11294	GATGGTGGGCATTTTTCGAATTGTCTCCGATGTATCACGCCACTTTTACTC-----	11341
Qy	304	LysAlaTyrGlySerAlaLysArgValAsnLeuGlu-----	315

```

Db 11342 -----TGGGATATGTCGACCTGGTGAATCTCTCCAGCGGTCCGGCTACCTGATTTC 11395
QY 316 ---LysGluPheProGlnSerTyrValGlnThrValGluAsnMetIleMetAlaLeuIle 334
Db 11396 GCTGAGCGTCTACCGCAA---TGGCAGAGGTTGTGGTCCAGAGATGAAGTGCTTCGT 11452
QY 335 SerIleSerLeuProAspTyrAsnThrProMetPheGlyAspSerTyrIleThrAspLys 354
Db 11453 AGCATCGACGATCCGGATGGTAGAATAGCTTCCTCAATGAT----- 11494
QY 355 AsnPheArgMetAlaGlnPheAlaSerTyrAlaArgValPheProAlaAsnGlnAlaIle 374
Db 11495 -----GCTGCATTCCGA-----ATTGCACCTGAATATGAAGATATT 11530
QY 375 LysTyrPheAla-----ThrAspGlyLysGlnGlyAlaProAsnPheLeuSerLys 392
Db 11531 CGAGCGTACCGCAGCGCTCGATATATCACCTCCAGCTCATGAACACCATCGCGCGG 11590
QY 393 AlaLeuSerAsnAla---GlyPheTyrThrPheArgSerGlyTyrAspLysAsnAlaThr 411
Db 11591 ATTATAATTCGGCTACAGTTATGCTGCTGACTTCTCCCGCATCGCGTGAAGCTATA 11650
QY 412 ValMetValLeuLysAlaSerPro-----ProGlyGluPheHisAlaGlnProAsp 428
Db 11651 TTGGATCTTGCAGAGTCGCTCCGACTATCAGCCGGGA-----CATGCGCATCGCGAC 11704
QY 429 AsnGlyThrPheGluLeuPheIleLysGlyArgAsnPheThrProAspAlaGlyValPhe 448
Db 11705 ACATTCAGTITTTAGTTAGCGCTTTTGAAACACGCTTAGTAGTCATATCCCGAATCTT 11764
QY 449 ValTyrSerGlyAspGluAlaIleMetLysLeuArgAsnTyrTyrArgGlnThrArgIle 468
Db 11765 CAGTACGCGTACGACTCTGAG-----CGACAAAGCGAGCGTGGTACTCGCGCG 11812
QY 469 HisSerThrLeuThrLeu 474
Db 11813 CATAATACGCTGGGANTG 11830

RESULT 14
AF498411
LOCUS AF498411
DEFINITION Pseudomonas aeruginosa serotype 019 putative O-antigen biosynthesis
ACCESSION AF498411
VERSION AC104730
KEYWORDS AF498411.1 GI:20559940
SOURCE Pseudomonas aeruginosa.
ORGANISM Pseudomonas aeruginosa.
REFERENCE 1 (bases 1 to 16720)
AUTHORS Raymond,C.K., Sims,E.H., Kas,A., Spencer,D.H., Kuttyavin,T.V.,
Ivey,R.G., Zhou,Y., Kaul,R., Clendenning,J.B. and Olson,M.V.
TITLE Genetic variation at the O-antigen biosynthetic locus in
Pseudomonas aeruginosa
JOURNAL J. Bacteriol. 184 (13), 3614-3622 (2002)
MEDLINE 22053227
PUBMED 12057956
REFERENCE 2 (bases 1 to 16720)
AUTHORS Raymond,C.K., Sims,E.H., Kas,A., Spencer,D.H., Kuttyavin,T.V.,
Ivey,R.G., Zhou,Y., Kaul,R., Clendenning,J.B. and Olson,M.V.
TITLE Direct Submission
JOURNAL Submitted (04-APR-2002) Genome Center, University of Washington,
Box 352145, Seattle, WA 98105-2145, USA
COMMENT On May 13, 2002 this sequence version replaced gi:17975259.
FEATURES
source
1..16720
/organism="Pseudomonas aeruginosa"
/serotype="019"
/db_xref="taxon:287"
<1..104
/note="ORF 1; C-terminal coding region of rpsA"
/codon_start=3
CDS

```

```

/transl_table=11
/product="RpsA"
/protein_id="AA027704.1"
/db_xref="GI:20559941"
/translation="EKDAMKELRKQEVESAGPTTIGDLIRAOQMNQG"
241..525
/note="ORF 2; himD/ihfB"
/codon_start=1
/transl_table=11
/product="HimD"
/protein_id="AA027705.1"
/db_xref="GI:20559942"
/translation="MTKSELIRIVTHOGLSAKOVLEIAKTMLEQMSQALATGDRIE
IRGGSFSLHYRAPRGRNPKTGESVRLDGKVFPHFKPKGKELDRVNEFE"
601..858
/note="ORF 3; hypothetical orfA"
/codon_start=1
/transl_table=11
/protein_id="AA027706.1"
/db_xref="GI:20559943"
/translation="WVSIVVVVFLVLENOGLQVAFGLQSPQWPLAVYLITAFVLGGL
LGLAIQLPSLAISRARASGLRAELEQARKEVDSLVRPSSS"
1114..2196
/note="ORF 4; wzz; similar to chain length determinant
protein"
/codon_start=1
/transl_table=11
/product="Wzz"
/protein_id="AA027707.1"
/db_xref="GI:20559944"
/translation="WTDGDSRQTSKRNEIDFIELIQLGWSOKWLVSLTFLGAVLAG
SVAFSLKPYERARVALLPSLSDVAGFIARGDSGLPPTFYQVYSVLRNQLSEGT
RKQFPREVYLSLEKSGIDGLYAFSDVSNRPSKDLPERYTVIVNRQPEQAAK
WAKQYADLAFLSDEMGRNQALEVSRNLIQDITLRQVVKERRNDRIVRLKEAL
QVARAVGEKPLIGAIQDOQLSALMEGSLMYMRGTALAEIKALEKRTSDPFPFA
LRGLQALALYESKOLNFRDVAVFQDGGVEIPDDPIKPKKALLALGILAGGVFVGL
LALLRLLDGRNGMADLKHPMTKDK"
2403..3488
/note="ORF 5; similar to UDP-N-acetylglucosamine
2-epimerase"
/codon_start=1
/transl_table=11
/protein_id="AA027708.1"
/db_xref="GI:20559945"
/translation="MAPVTALADDDRFEARVCVTGQHREMLDQVNLNFKIIVDFDLN
INRSGDLTDVTVAIQGKQVFSQFKPDIVLVHGDITATLAASLAAYHQIPVAHVE
AGLRTGNNYSPWPERGKRLTGALACLFAPETGRDNLREGISAERIRVTGNTVID
ALLNVQRLQODMELASAAAPTEFLDSSRRFILVTHRRSEFGGGERICQALGVVA
RNHPEDIVYPVHLNPNVREPUNRLISDVQNIHLIEPLDYLFPVHLSMAHVILTDSG
GICEEAPSLGKPLVMRETERPEAVAAGTVKLVGTEIKNIVNELERLLVDSAYRKM
SFAHPYGGGWASKRIIDSLDLAKRGE"
3569..4837
/note="ORF 6; similar to UDP-glucose/GDP-mannose
dehydrogenase"
/codon_start=1
/transl_table=11
/protein_id="AA027709.1"
/db_xref="GI:20559946"
/translation="MSFTISVIGLYIGLPTAAVFASRRKQVIGVDVNSKAVETINQ
KAIHIVPEOLDMLVHAATVYRAVSVPEPADAFLIAVPTPFKEQHPDLRYIESAC
KALAPVAKAGDVLILETSVPGATQMSAWLSAERPDLTFQSGHSDADRIAHCPER
VLPHVRLVQNDRVIGGMTATCSAAASLYKIFVEGECVITNARTAMCKLITNSF
ADTNIAFANELISICLDLINWELIRLANRHPVNILOPQGVGGHGIADVDPWIFS
KTPGEARLIRTAREVNDKGLVNLNKIEMVAEFLQANPRTAKDVTACIACFAKFPD
ILLDRSPASVIAQNVKMSHPGTVLWPEPIAELPEKLAERLDLVSINSALERSEVIV
LLVDHKEFKFSPCDLGAYKYVIDTRGVNA"
4915..6177
/note="ORF 7; similar to Polysaccharide biosynthesis
protein; potential multiple membrane spanning domains."
/codon_start=1
/transl_table=11
/protein_id="AA027710.1"
/db_xref="GI:20559947"

```

```
/translation="WLKLNENSVNLSVALOAGNALLPLLAPPLFLGVLEQGAFAF
LVAALAFYVLTVCLYSFDSISGVQIVDAKEEREVEAKFPFNFAARVGLLIPSSVF
LVGGTLFWGADVLVLLVFLVGLQILQSNFYFOAVENNNLLATFVSSRVAVIAV
YVGHKGDTLWLASILVLAGSVGGVAGVGLLTFRGLSEKRRITGDVGVSSLLSGRH
LFLGNLSYTRFRGNALVLLAGVNSAAVVALSEKIIFKQALARPQLNQLPKVVK
AWASLERNNAPFLVKNNTVAQVLLMVFLLPVFVFLFKVGLVGIWPGSEKVF
FLVIMSPAVLIPGVANSMFGVAGLSLGSQTYFAVSVISIGVCAFLISLVMSYLLSDT
GAAMTFVLAELLFLFAFLCKVRGRPHG"
```

CDS

```
6170..7504
/Note="ORF_8; similar to u-PAR/ly-6 domain; potential
multiple membrane spanning domains."
```

```
/codon_start=1
/transl_table=11
/protein_id="AAM27711.1"
/db_xref="GI:20559948"
/translation="MDEFFYLKLINIFLSPFLGAWCARKIFSVLNPVIFSLWPF
FYTPFLVAPEAPVPAIILYIFCFVSLSPFPGWAFSPNKNKSNACSYFDV
PLMWLAFCGSAIMIFWGLQOQISLEQLNPNVSIIGIYAGKRYSEIVSVFVS
PVLQCSYTVAVFGLIYGAARSEARGLKLPFTLPALLVMLQSAKGLFPFISFLF
FGMLSVRYKNVTLVGKDFRLLYGLLILPLIFSLSRGIYQDDTDALMNRLL
RYLVTYSVHLPAESDMFSEYFCESLMHYROESSTLGFYTPMSFFOLAGDSDVAM
GYIDFFGYEYKNTLYTVRGLITDFLGLSLFFAFILGLFCMLCWRLLCRNSGA
FAILFFCYFAISYQTYIISLTLTLPLFPFVIFQWFMGLLKFYFTRSKV"
7501..8343
/Note="ORF_9; similar to Glycosyl transferase"
```

CDS

```
/codon_start=1
/transl_table=11
/protein_id="AAM27712.1"
/db_xref="GI:20559949"
/translation="MKICLVPCVYNADHDVRFLLDSVNTAFSCFLGVLEVLSDNST
LSSVITFDARYSNFYRQLNSNIGYFAPAFALSTLPEIDSYDFVMVNDVLVSK
DFFIALSPGKQVGLIAPGIFSDKGDGLNPKIMRPSLRKMKFMVMVCSSTALFR
WYSLVABRETRACTQRIAKKGAIQGPAMYGAHGSFIIFLQLYEKGHVLVPR
FLFGEFVFAEQRLNNMLIQHPDVRVSDKEHASTSKIKLKFICPEHRKSYEYLLSK
FPDD"
```

CDS

```
8457..10625
/Note="ORF_10; similar to Zinc-binding dehydrogenases"
/codon_start=1
/transl_table=11
/protein_id="AAM27713.1"
/db_xref="GI:20559950"
/translation="MSFCIEGAMKQVLQDMAGKGTSVTVAPAPQISRDIHINSTVSL
ISAGTERMLVFGKASYLDKQAPQKVMVLEKIDGLMTTIRVAVQSOKLAQPLVLA
YCNVGVSEISGSGVGVKGVDRVNSGPHADLVKYSKNCILCARIIPGVDDSEAAFPVGA
SIGLOGIPLAOTLGEVFTVGVGLITVQLLRAHGORVLAIDFDSKLELQRYG
AQTCPNGKGEDVPAAGMFAFSKGVGVGVIITASTKSDPTVQAARMKRGRVILVGV
VGLNLEADFEKELTQVQSGPGRYDPEYEKGGDPLGFWKMTBORNPEAVLDM
LDSQLEVKTLITHRKPEDAPAYSTLTQDNSGLMLLYASDTSRVRVHVPLSDM
KGFNAQKPVGVFTAGNYASRILIPAFKTAGAQHTLVTSNGINSVHGEKAGFSQAS
TNIDAVLNERMTAIIVTRHDTARLVVDALRTGNVFEVEXPLAITHAELEDVQAVH
ALHTQGGSPOLMVFNRREAPHIOKMKGILLASVREPKSFIMTMNAGALPUGHWTQDI
AVGGRIIGEACHFIDLMRYLAGSRIVSVQARRMGDSNAIDTSEKASITLGFEGDSP
GTILYLANGASSFPKRIEYVFTAGRVQLQDNFRKLKGYGPGFNKNLWKQDKGQRC
VAAFLDSIEKGEPTIACEBIEFVTRVSIIEAAEILRNQMD"
```

CDS

```
10634..12286
/Note="ORF_11; similar to Prenyltransferase and squalene
oxidase"
/codon_start=1
/transl_table=11
/protein_id="AAM27714.1"
/db_xref="GI:20559951"
/translation="NALQKLVRLYHTRLYRLKQIFRVVYFAQVRVHYRNVARR
WRDWSAPSRNLSNTDAVEFTFLGITGVQVCPEDWQADRSKMLYNLHYLDDLNAR
DIGSQGLADLQSIQANPPVSGEGWEPYPLSRIUNLVKARHBERSTFLADSL
AVQADLVQOVBHYHILGNHLFANGRLVFAAGLAYLSGAMADRWLAKRLRDLDELPEQF
LNDGHFELSPMHTATLLMDCDLVNLSRGLPDLAERLPQWRVGVVQGLKWSMQ
HPDGRISFNDAFPAEYEDIAAYAKELDISPAHENHLAAIYNSTATGYAAVLPA
GVKALIDLAKVGPDPQGHADLTLSFELSVFKRILVNSGTSQYCDSDSERQORQTA
RNTVTLGLGSDSVSWAGVRVRAAVVIERVEPEPEITCIOASHDGVRLVKGLRHR
RTWTFTRSMELFDSITGDOVVATSRFFPFSPEVKYMKADGFVADMGAGKKYSIDPSG
SDVRLIASSWHPFGCSLANQCLVATFSGDSLTITRICHGGI"
```

CDS

```
12355..13467
/Note="ORF_12; similar to Glycosyl transferases group 1"
```

```
/codon_start=1
/transl_table=11
/protein_id="AAM27715.1"
/db_xref="GI:20559952"
/translation="MSALLGALPENSQLEVIITPNRYSSFSLEAPLEKHSRLIHR
IPITQHNSGMLDQRAFFSYLGLVITRNKDYDVVTATSSRLMTASIGALVARQLRC
PLYLDIRFDVTIKDVLPSRFPVKPIFSMERITVTRATRVNLVSGALFYPNSR
YPGDFVVTNGIDIDEFVKVQTSCEERESSLYTLVYAGNMGGQGLHLIIPQIARHF
DGHRLFCIGDGRKQLQDAIBEAGCTNVTITSPRSELIAYAKADVLFLHNDY
DAFRVLPSKIFIEYAAKGLPWAGVAGSAGFIQDITSNAAVFRPCNVEDALKAFSSL
SMVTKPRKEFVERFAREKIMKALADIIIFVVRGEV"
```

CDS

```
13464..14417
/Note="ORF_13; similar to NAD dependent
epimerase/dehydratase family"
```

```
/codon_start=1
/transl_table=11
/protein_id="AAM27716.1"
/db_xref="GI:20559953"
/translation="MRVPLTIGITGVGLATSKRLIQEGHCYVAAVRDANPALAORVEQ
VVFNLVKHDWASSLEGOEVVHCAARVHVMMKASRDPLSEFRVNVREGTINLARQAA
KAGYKRFVFISSIKVNGESTRLGVPTADDIAMPDLPYGVSKLEAEQKLLSLAARVDM
```

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
0.0015	170.50	88	
Percent Similarity:	36.61%	Conservative:	46
Best Local Similarity:	24.04%	Mismatches:	151
Query Match:	4.88%	Indels:	81
DB:	1	Gaps:	18

US-09-802-285a-2 (1-659) x AP498411 (1-16720)

QY	144	TrpGlnAlaMetAlaLeuValTyrHisAlaThrGlyAspGluLysTyrAlaArgGluTyrP	163
DB	10871	TGGCAAGCT-----GATGAGCAGTAATAATATG	10900

QY	164	ValTyr-----GlnTyrSerAspTrpAlaArgLysAsnProLeuGlyLeuSerGlnAsp	181
DB	10901	CTCTATAATCTGCACATAITTTGGATGATCTGAACGCTAGGACATTGGC---AGTCAGCCT	10957

QY	182	Asn-----AspLysPheVal-----	186
DB	10958	GGGCTAGCCGATAAACTAATTCAAAGTTGGATCCAGGCTAATCCACCGGTATCCGGTGAA	11017

QY	187	---TrpArgProLeuGluValSerAspArgValGlnSerLeuProThrPheSerLeu	205
DB	11018	GTTTGGGACCTTATCTCTTTCCTTGGTATCGTAAATCTTGTGAATGGCTTGTCTGC	11077

QY	206	PheValAsnSerProAlaPheThrProAlaPheLeuMetGluPheLeuAsnSerTyrHis	225
DB	11078	CATGATGACGGTTCGACCTTT-----CTGCGGACAGTCTAGCT	11116

QY	226	GlnGlnAlaAspTyrLeu-----SerThrHisTyrAlaGluGlnGlyAsnHisArgLeu	243
DB	11117	GTACAAGCGGACGCGCTCGTGCAACAGTAGAATAATCATATCTTGGAAATCATCTATTTC	11176

QY	244	PheGluAlaGlnArgAsnLeuPheAlaGlyValSerPheProGluPheLysAspSerPro	263
DB	11177	GCTAATGGGACGCGTGTGGTGTTCGCGC---GCTTATCTCAGGCGCGATGGCTGAT	11233

QY	264	ArgTrpArgGlnThrGlyIleSerValLeuAsnThrGluIleLysGlnValTyrAla	283
DB	11234	CCTTGGCTTGGCAAGGGTTACGGATACTTTCAGGAAGAGCTCTCTGAGCAATTTCTGAAT	11293

QY	284	AspGlyMetGlnPheGluLeuSerProIleTyrHisValAlaAlaIleAspIlePheLeu	303
DB	11294	GATGTGGGCATTTCCAAATGTCTCCGATGATATCAGCCACTTTTACTC-----	11341

QY	304	LysAlaTyrGlySerAlaLysArgValAsnLeuGlu-----	315
DB	11342	-----TGGGATATGTGGACCTGTGGTGTTCGCGC---GCTTATCTCAGGCGCGATGGCTGAT	11395

QY	316	---LysGluPheProGlnSerTyrValGlnThrValGluAsnMetIleMetAlaLeuIle	334
----	-----	--	-----

```
Db 11396 GCTGAGGCTACCGCAA---TGCGAGAGTGTGGTCCAGGATGAAGTCTCGT 11452
QY 335 SerIleSerLeuProAspTyrAsnThrProMetPheGlyAspSerTrpIleThrAspLys 354
Db 11453 AGCATCAGCATCCGATGGTAGAATTAGCTTCTCAATGAT----- 11494
QY 355 AsnPheArgMetAlaGlnPheAlaSerTrpAlaArgValPheProAlaAsnGlnAlaIle 374
Db 11495 -----GCTGCATTCGGA-----ATTGCACCTGAATATGAAGATATT 11530
QY 375 LysTyrPheAla-----ThrAspGlyLysGlnGlyLysAlaProAsnPheLeuSerLys 392
Db 11531 GCAGCGTAGCCCAAGCGCTCGATATATCACCTCCAGCTCATGAAACACACCTGGCGGC 11590
QY 393 AlaLeuSerAsnAla---GlyPheTyrThrPheArgSerGlyTrpAspLysAsnAlaThr 411
Db 11591 ATTATTAANTTCGGCTACAGGTTATGCTGTGTTACTTCTCCGATGCGGTGAAGCTATA 11650
QY 412 ValMetValLeuLysAlaSerPro-----ProGlyGluPheHisAlaGlnProAsp 428
Db 11651 TTGGATCTGCAAGAGTCGTCGCCGACTATCAGCCGGA-----CATGCGCATGCCGAC 11704
QY 429 AsnGlyThrPheGluLeuPheIleLysGlyArgAsnPheThrProAspAlaGlyValPhe 448
Db 11705 ACATTGAGTTTGGTAAAGCGCTTTGGAAACGCTTAGTAGTCAATCCGCACTTCT 11764
QY 449 ValTyrSerGlyAspGluAlaIleMetLysLeuArgAsnTrpTyrArgGlnThrArgIle 468
Db 11765 CAGTAGCGGTGATGACTCTGAG-----CGCAAAGGCGCGTGTACTGCGGCG 11812
QY 469 HisSerThrLeuThrLeu 474
Db 11813 CATAATACGGTGGGATTG 11830

RESULT 15
AL646082/c
LOCUS
DEFINITION Ralstonia solanacearum GMI1000 megaplasmid, complete sequence;
segment 7/11.
ACCESSION AL646082 AL646053
VERSION AL646082.1 GI:17431428
KEYWORDS
SOURCE
ORGANISM Ralstonia solanacearum.
Ralstonia solanacearum
Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
Ralstonia.
1 (bases 1 to 205050)
Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,
Arlat,M., Billault,A., Brottier,P., Camus,J.C., Cattolico,L.,
Chandler,M., Choise,N., Claudel-Renard,C., Cunnac,S., Denange,N.,
Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schlex,I.,
Siguier,P., Thebault,P., Whalen,M., Wincker,P., Levy,M.,
Weissenbach,J. and Boucher,C.A.
Genome sequence of the plant pathogen Ralstonia solanacearum
Unpublished
2 (bases 1 to 205050)
Boucher,C.A.
Direct Submission
Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
Moleculaire des Interactions Plantes-Microorganismes INRA-CNRS,
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, IMGM CNRS
118 Route de Marbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
F31326 Castanet-Tolosan Cedex, Laboratoire de Genetique Cellulaire
INRA, BP27, F31326 Castanet-Tolosan Cedex
Christian.Boucher@toulouse.inra.fr
http://sequence.toulouse.inra.fr/R.solanacearum.html.
Location/Qualifiers
1..205050

/organism="Ralstonia solanacearum"
/strain="GMI1000"
/db_xref="taxon:305"
/plasmid="megaplasmid"
complement(486..845)
/gene="RSP0956"
/note="RSP0956"
/note="RSP0956"
complement(486..845)
/gene="RSP0956"
/function="miscellaneous; unknown"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Framed"
/evidence=not experimental
/codon_start=1
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAD18107.1"
/db_xref="GI:17431429"
/translation="MSEKLGAEELMPENWRDDLEALDLBIVRPAICQVKLLPEPGVIQ
RVLDNDARDCHIGNOTAFASRLGLLFLHFQMKDLAEAYGPSDATEIVRRVVAHLOPR
VSDQLGSLPNPGALSAV"
1409..1714
/gene="RSP0957"
/note="RSP0957"
1409..1714
/gene="RSP0957"
/function="miscellaneous; unknown"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Framed"
/evidence=not experimental
/codon_start=1
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAD18108.1"
/db_xref="GI:17431430"
/translation="MEHDKYLHAGISAVIASGVTTLAVDSPNRFYGLGAGMAVGLTK
EIVDRKTNRGDRFSDLLADLAGALIGAYADSLLRPAVFKPTGYAYGVLNVAFD"
complement(1869..2540)
/gene="RSP0958"
/note="RSP0958"
complement(1869..2540)
/gene="RSP0958"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD18109.1"
/db_xref="GI:17431431"
/translation="MPLDQPLGLHARNIPGATGIFHEYQLDFCCGQHSRLDAAQAK
GIDAAPIAARLQALQAEATPGADVWSAVPSMLIDHILRPHRHRQLPELIRLAR
RVEHVHGRPCPCVGLSELLAMQELSHMKKEQILFPMIARGHGLRAGGPIAVMR
MEHDHGALQRLMTLTNDITPPRAACTTWRALYGLSVFREDLMEHLHLENNILFEG
AAVAD"
2706..4238
/gene="RSP0959"
/note="RSP0959"
2706..4238
/gene="RSP0959"
/function="miscellaneous; not classified regulator"
/note="Product confidence : probable
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Homology
```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 28, 2003, 20:50:03 ; Search time 39 Seconds
(without alignments)
2006.736 Million cell updates/sec

Title: US-09-802-285A-2

Perfect score: 3494

Sequence: 1 MTTKFKRIIVFAVIALSSG.....KGLNLTTLTKQQLVLP 659

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Lasting first 45 summaries

Database : Published Applications AA.*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	128	3.7	25	10	US-09-802-285-3
2	128	3.7	25	15	US-10-291-337-3
3	116	3.3	4563	11	US-09-802-640-32
4	115.5	3.3	1331	10	US-09-801-368-370
5	113.5	3.2	1398	14	US-10-090-624-6
6	110.5	3.2	563	9	US-09-815-242-5665
7	110.5	3.2	578	9	US-09-815-242-12206
8	110	3.1	4563	10	US-09-870-759-128
9	107.5	3.1	1336	11	US-09-934-455-22
10	107.5	3.1	1336	15	US-10-278-173-128
11	106	3.0	1375	9	US-09-740-274-4
12	105.5	3.0	1167	9	US-09-815-242-11522
13	105	3.0	965	11	US-09-842-484A-2
14	105	3.0	965	15	US-10-184-485-3
15	105	3.0	1135	11	US-09-759-130B-42

16	104	3.0	1849	15	US-10-242-056-49	Sequence 49, Appli
17	104	3.0	2516	9	US-09-817-514A-2	Sequence 2, Appli
18	104	3.0	2516	15	US-10-242-056-47	Sequence 47, Appli
19	103.5	3.0	1183	10	US-09-870-759-45	Sequence 45, Appli
20	103.5	3.0	1599	14	US-10-092-880-9	Sequence 9, Appli
21	103	2.9	920	9	US-09-815-242-12995	Sequence 12995, A
22	103	2.9	920	9	US-09-815-242-13148	Sequence 13148, A
23	102	2.9	26926	10	US-09-759-508B-2	Sequence 2, Appli
24	101.5	2.9	943	9	US-09-969-362-5	Sequence 5, Appli
25	101.5	2.9	964	15	US-10-100-049-74	Sequence 74, Appli
26	101.5	2.9	965	11	US-08-842-484A-4	Sequence 4, Appli
27	101.5	2.9	965	15	US-10-217-613-3	Sequence 3, Appli
28	101.5	2.9	965	15	US-10-217-613-8	Sequence 8, Appli
29	101	2.9	917	9	US-09-815-242-5503	Sequence 5603, Ap
30	101	2.9	920	9	US-09-815-242-12181	Sequence 12181, A
31	101	2.9	972	15	US-10-217-613-9	Sequence 9, Appli
32	100.5	2.9	664	9	US-09-815-242-12577	Sequence 12577, A
33	100.5	2.9	691	9	US-09-815-242-5223	Sequence 5223, Ap
34	99.5	2.8	690	9	US-09-815-242-12460	Sequence 12460, A
35	99.5	2.8	703	9	US-09-815-242-5610	Sequence 5610, Ap
36	99.5	2.8	804	14	US-10-011-588-37	Sequence 37, Appli
37	99	2.8	633	9	US-09-871-615-5	Sequence 5, Appli
38	99	2.8	1477	14	US-10-092-880-4	Sequence 4, Appli
39	98.5	2.8	556	9	US-09-815-242-5527	Sequence 5527, Ap
40	98.5	2.8	565	9	US-09-815-242-12352	Sequence 12352, A
41	98.5	2.8	565	9	US-09-815-242-12931	Sequence 12931, A
42	98	2.8	615	10	US-09-908-193-26	Sequence 26, Appli
43	98	2.8	944	15	US-10-213-990-27	Sequence 27, Appli
44	98	2.8	1187	9	US-09-768-436-4	Sequence 4, Appli
45	97.5	2.8	515	9	US-09-815-242-4928	Sequence 4928, Ap

ALIGNMENTS

RESULT 1

US-09-802-285-3
; Sequence 3, Application US/09802285
; Patent No. US20020122793A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Dongfang
; APPLICANT: Pojasek, Kevin
; APPLICANT: Shriner, Zachary
; APPLICANT: Holley, Kristine
; APPLICANT: El-Shabrawi, Yusuf
; APPLICANT: Venkataraman, Ganesh
; APPLICANT: Sasisekharan, Ram
; TITLE OF INVENTION: Heparinase III and Uses Thereof
; FILE REFERENCE: M0656/7063HCL
; CURRENT APPLICATION NUMBER: US/09/802,285
; CURRENT FILING DATE: 2001-03-08
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 3
; LENGTH: 25
; TYPE: PXT
; ORGANISM: Flavobacterium heparinum
US-09-802-285-3

Query Match 3.7%; Score 128; DB 10; Length 25;
Best Local Similarity 100.0%; Pred. No. 3.8e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 280 QYADGMQFELSPIYHVAIDIFLK 304

Db 1 QYADGMQFELSPIYHVAIDIFLK 25

RESULT 2

US-10-291-337-3
; Sequence 3, Application US/10291337
; Publication No. US20030099628A1
; GENERAL INFORMATION:

Db 123 GVKFFSTTPYCKMR-----SDSDELAWEIATERFKWQSMRLARVLKGDIVKGEKTRIAN 177
Qy 163 WYQYSDWARKNPLGLSQNDKFWVRPLE-----VSDRVQSL-----PPTFSLFVNSPA 211
Db 178 QV-----KKP-GLINKELSDIEIWLKAWLNGRMTQMEQSLTYLRDSSDSVFEIIMK 228
Qy 212 F-----TPAFLMEFLNSVHQOQADY---LSTHVAEQ--GNHRLFEAQ-----247
Db 229 FOIPOGKILSDALEAILQDLNMRHVSVYWPNLKMKYKPIINTAEFTARIIDVMSW 288
Qy 248 -----RNLFAGVSF-----PEFKDSPRWRTGTSVLNTEIKKQVADG 285
Db 289 LNFKNLTLRRQELDDWINRFSPISSSONCQEDFGVPQW-NCKMKILAEQLMKEKNIES 347
Qy 286 MQFELSPIYHVAADIFLKAYGSAKRVNLEK-----EFP-----Q 320
Db 348 I-FQKKIYFPLSPMFKLHFIYVRETLTKMKIKYYPYERLSLLAFVYLKIVELTRL 406
Qy 321 SYVQTVENMIMALISISLPDYNT-----PMFGDSWITDK 354
Db 407 SYARKLKPTMMIDQIDDFAFIRLSVLQKYTLTKYCSNLPFDVDEPTFENTVI---463
Qy 355 NFRMAQFASWARVFPANQAIKY-----FATGKQG-KAPNFLSK---ALSNAQFYTF 402
Db 464 -----EAIYLFLLNLKLISSKQNFKAPDLLLKYYWDHLKNTGHIYI- 505
Qy 403 RSGWKNATVMVLKASPPGFEHAQDNGTFEL---FIKGRNTTPDAGVFVYSGDEAIMKL 459
Db 506 -----NGAETVI-----PNEFLKTLRLVHKLOFYLLKQONFPPTFA---NASEA---547
Qy 460 RNWYQTRIHS'TLTDNQNMWITKARONKWTGNNLDVLTYPNSVYNLDHQRSVLFINK 519
Db 548 EKW-----LSSIFENLGMKKLRNF-----SNILVAFQNSAVYQINHAQV---K 592
Qy 520 K-----YFLVIDRAIGATGNLGVHW-----QLKEDSNVPFDKTRN-----VYTYRDG 564
Db 593 KLDKADHYFLVYS--GNTPESSGVYFAAPPELLGCDNDTILRLNKSIGCDLVKPLDIG 649
Qy 565 NNLMQSLNADRTSLN-----EEGKVSY---VYNKELKRFAPFVKPKKNAQTGNF---613
Db 650 NNLANVYDITTKETDLNLSVKGEDSKGIPYRVWANSSSDLDRAHQSKKNFSTDPDQ 709
Qy 614 -----VSIVVP-----YDGQKAPESIRENK-----GNDFEKGKL- 643
Db 710 HLDKKNVEFEVALSSLCALVLYPGEVVDG---PVYKLPGNLNFASNEMDLKIG 766
Qy 644 --NLTLTIN 650
Db 767 NENTLILN 775

RESULT 5

US-10-090-624-6

; Sequence 6, Application US/10090624

; Publication No. US20020132335A1

; GENERAL INFORMATION:

; APPLICANT: TAKAKURA, Hikaru

; APPLICANT: MORISHITA, Mio

; APPLICANT: SHIMOJO, Tomoko

; APPLICANT: ASADA, Kiyozo

; APPLICANT: KATO, Ikunoshin

; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE

; FILE REFERENCE: TAKAKURA-6

; CURRENT APPLICATION NUMBER: US/10/090,624

; CURRENT FILING DATE: 2002-03-06

; PRIOR APPLICATION NUMBER: 09/445,472

; PRIOR FILING DATE: 1999-12-06

; PRIOR APPLICATION NUMBER: 151969/1997

; PRIOR FILING DATE: 1997-06-10

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: Patent in version 3.0

; SEQ ID NO 6

; LENGTH: 1398

; TYPE: PRT

; ORGANISM: Pyrococcus furiosus

; US-10-090-624-6

Query Match 3.2%; Score 113.5; DB 14; Length 1398;

Best Local Similarity 18.5%; Pred. No. 0.47;

Matches 118; Conservative 83; Mismatches 233; Indels 203; Gaps 31;

Qy 134 VRMQLHRVKKWQAMALVYHATGDEKYAREWVYQYSDWARKNPLG--LSQNDKFWVRPLE 191
Db 703 VEWHIKYV-----GDTEYRTFEIYATEPWIKPFPVSGSVILENTEFV---LR 746
Qy 192 VSDRVQSIPPTF---SLFVNSPAFTPALMEFLNS-----YHQOADYLSHY---AQ 238
Db 747 VKYDVEGLEPGLYVGRITIDDP-TPVIEDEILNTIVPEKFTPENNYTLTWYDINGPEM 805
Qy 239 GNHRLFEAQRLNLFAGVSPEFKDSPRWRTGTSVLNTEIK--KQVYADGMQ-----287
Db 806 VTHHFTTPEGVDVLYAMTTYWDYGLYRPGDMFVFPYQLDYLPAAVSNPMPGNWELVWG 865
Qy 288 FELSPIYHVAADIFLKAYG-----SARKVNLEKEFPQSYVQTVENMI---MALIS 335
Db 866 FNFAPLYESG---FLVRIYGVETPSVMYINRTYLTDTNTEFSIBFNITNIYAPINATLIP 922
Qy 336 ISLPDYNTPM--FGDSWITDKNFRMAQFASWARVFPANQAIKYFATDGKQKQKAPNLSKA 393
Db 923 IGLGTYNASVESVGGEFFIKGIEVPEGTAEULKIRIGNPSV-----PN-----965
Qy 394 LSNAGFYFRSGWKNATVMVLKASPPGFEHA---QPDNGTFELFKGRNFTPDAGVFYV 450
Db 966 -SDLLDLXYL---DSKGNLVALDGNPTAEVEVVVEPKPGVYSIVVHGYSVVDENGNTPT 1020
Qy 451 SGDEAIMKLRNRYOTRIHSTLTLDNQNMWITKARONKWTGNNLDVLTYPNSVYNLDH 510
Db 1021 TTFDLVWQ-----MTLDNGNIKLDK---DSIILGSNBSVVVTANITTI-DRDH 1063
Qy 511 QRSV-----LFINKKY-----FVIDRA-----IGEATGNLGVHWQLKEDSN 547
Db 1064 PFGVSGIIEIRDNEVYQDTNTSIAKIPITLVIDRADFAVGLTPAEGVLG-----1113
Qy 548 PVFDTKRVVT-----TYRDGNNLMITQSLNADRTSLNBEKGVSVYV-----590
Db 1114 -----EARNYTLIVKHALTLEPVENATV--IIGNVTYLTDENGVTFTFYAFTKLGSDEI 1165
Qy 591 -----NKELKRPAPVFEKPKKNAQTGNFVSVIYDQ---622
Db 1166 TVIVKKNFNTLEKTFQITVSEPEITBEDINEPKLAMSPEANA---TIVSVEMESEGCV 1222
Qy 623 OKAPEISIRENKGNDFEKGKLNLTITNGKQQLVLP 659
Db 1223 KATVVEITIN-----GTANETATI-----VVPVP 1247

RESULT 6

US-09-815-242-5665

; Sequence 5665, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

```
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5665
; LENGTH: 563
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5665

Query Match      3.2%; Score 110.5; DB 9; Length 563;
Best Local Similarity 18.1%; Pred. No. 0.21;
Matches 105; Conservative 94; Mismatches 211; Indels 171; Gaps 26;

QY 91 IDKVTREMAKALVHQFQPHKGYGYDYGKDINQMWPVKDNEVRWQLHRVKKWQAMA-L 149
DB 8 IERILNEFRENLQVPPQN-----HWSIQLESELTLTGQFAYWSAMGHH 54

QY 150 VYHATG---DKYAREWVYQSDWARKNPLGLSDNDKFWWRPLEVSRVOSLPPTFSLF 206
DB 55 MYHSEVWLIDGSKSLTYKEAIAIRILQHMAQSDNQTAVQOHM----- 98

QY 207 VNSPAFTAFALMEFLNSVHQADYLSTHYAEQGNHLEFAEQNLFAGVSE-PEEKDSPRW 265
DB 99 -----AQIMSDIDNSIHRTRYLQSNITIDYAEDRVIVSQSLYLGHPPHPTPKSASGF 151

QY 266 RQTGIS-----VLNTEIKQVYADQMOPFELSPI-YHVAADID-----FL 303
DB 152 SEADLEKYAPBCHTSFQLHYLAHQDVLLTRYVEGKEDQVEKVLQADIDISEIPKDFI 211

QY 304 KAYSARKVNLEKEFPQSYQTVENMIMALLISISLDPYNTPMFGDS----- 349
DB 212 LLPHPYQINVLQOHQP-YLQYSEQGLIKGLGVS-----GDSVYPTSSVRTVFSKA 261

QY 350 -----WITDKNF-RMAQFASWARVFPANOAI-----KYFATDGKOGKAP 387
DB 262 LNIYKLEPIHVKINFIRTNLEQIERIDIAQVIASVKDEVEPHFKLMFEEGVRLALP 321

QY 388 NFLSKA-----LSNAGFYTFRSG-----WDKNATVMVLKASPPGFHQAQDNGTFEL- 434
DB 322 NPLGQTVPEPMDLLTNSAMIV-REGIPNYHADKDIHVLA-----SLFETWPDSPSLS 374

QY 435 -FIKGRNFTPDAGVVFVSGDEAIMKLRNWRQTRI-----HSTLTL-----D 475
DB 375 QVIEQSGLAPEAWLECYL-DRTLPLILKLYSNTGISLEAHVQNSLIELKDGIPDVCVFRD 433

QY 476 NONNVITKARQNKWETGNNDLVLTNPSPNLDHORSVLFIKKYFLVIDR-----AI 529
DB 434 LEGICLSRTIATEKQLVFN--VVAASSP-----VVAHDEAHRLLKYYVNVNHLGHLVSTI 487

QY 530 GEATGNLGVHWQL-----KE-DSNPVFDKTKNRVYTT 560
DB 488 GKATKNEVVLQVAHRLMTWKKEYANNAVDCVEDIYQT 528

RESULT 7
US-09-815-242-12206
; Sequence 12206, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
```

```
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12206
; LENGTH: 578
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12206

Query Match      3.2%; Score 110.5; DB 9; Length 578;
Best Local Similarity 18.1%; Pred. No. 0.22;
Matches 105; Conservative 94; Mismatches 211; Indels 171; Gaps 26;

QY 91 IDKVTREMAKALVHQFQPHKGYGYDYGKDINQMWPVKDNEVRWQLHRVKKWQAMA-L 149
DB 15 IERILNEFRENLQVPPQN-----HWSIQLESELTLTGQFAYWSAMGHH 61

QY 150 VYHATG---DKYAREWVYQSDWARKNPLGLSDNDKFWWRPLEVSRVOSLPPTFSLF 206
DB 62 MYHSEVWLIDGSKSLTYKEAIAIRILQHMAQSDNQTAVQOHM----- 105

QY 207 VNSPAFTAFALMEFLNSVHQADYLSTHYAEQGNHLEFAEQNLFAGVSE-PEEKDSPRW 265
DB 106 -----AQIMSDIDNSIHRTRYLQSNITIDYAEDRVIVSQSLYLGHPPHPTPKSASGF 158

QY 266 RQTGIS-----VLNTEIKQVYADQMOPFELSPI-YHVAADID-----FL 303
DB 159 SEADLEKYAPBCHTSFQLHYLAHQDVLLTRYVEGKEDQVEKVLQADIDISEIPKDFI 218

QY 304 KAYSARKVNLEKEFPQSYQTVENMIMALLISISLDPYNTPMFGDS----- 349
DB 219 LLPHPYQINVLQOHQP-YLQYSEQGLIKGLGVS-----GDSVYPTSSVRTVFSKA 268

QY 350 -----WITDKNF-RMAQFASWARVFPANOAI-----KYFATDGKOGKAP 387
DB 269 LNIYKLEPIHVKINFIRTNLEQIERIDIAQVIASVKDEVEPHFKLMFEEGVRLALP 328

QY 388 NFLSKA-----LSNAGFYTFRSG-----WDKNATVMVLKASPPGFHQAQDNGTFEL- 434
DB 329 NPLGQTVPEPMDLLTNSAMIV-REGIPNYHADKDIHVLA-----SLFETWPDSPSLS 381

QY 435 -FIKGRNFTPDAGVVFVSGDEAIMKLRNWRQTRI-----HSTLTL-----D 475
DB 382 QVIEQSGLAPEAWLECYL-DRTLPLILKLYSNTGISLEAHVQNSLIELKDGIPDVCVFRD 440

QY 476 NONNVITKARQNKWETGNNDLVLTNPSPNLDHORSVLFIKKYFLVIDR-----AI 529
DB 441 LEGICLSRTIATEKQLVFN--VVAASSP-----VVAHDEAHRLLKYYVNVNHLGHLVSTI 494

QY 530 GEATGNLGVHWQL-----KE-DSNPVFDKTKNRVYTT 560
```


Matches	147;	Conservative	89;	Mismatches	252;	Indels	306;	Gaps	41;
Qy	16	ALSSGNILAQSSITRKDFD	---	---	HINLXSGLEKVNKAAGNYDDAAKALL	64			
Db	641	ALSYALLLTNKSPPRVVYGMFTDDGGYMAKHTINVEAETL	---	---	---	---	---	---	---
Qy	65	AYYEKSKAREPDESNAEKPADIR	---	---	QPIDKVTR	---	---	---	---
Db	689	KYVSGGQAMRQOVGNSEBIITSRYGKGALKATDGTDRTRTSGVAVIEGNNPSLRILKAS	748						
Qy	100	DKALVHOFQPHKGYG	---	---	FDYKGDINNMQWPFK	---	---	DNE	---
Db	749	DRVVVNAAHKNQAYRPLLLTTDNG	---	---	IKAYHSDQEAAGLVRYTNDGE	---	---	---	---
Qy	146	AMALVYHATGDEKYAREWVQY	---	---	SDWARKNPLGLSQDNDFKVMRPLEVSDRVQSLPPTFS	203			
Db	797	---	---	---	LIFTAADIKGYANPQVSGYLGVW	---	---	PVGAAADQD	---
Qy	205	LFVNSPAPTFAFLMEFLNSVHQADYLSHTHYAEOG	---	---	NHRLFEAQRNLFAGVSPFEKDS	262			
Db	843	---	---	---	KSVEHNA	---	---	ALDSRVMEFGSNFOAFATKBEYTNVIA	---
Qy	263	PRWRQTGISVLNTEIKKQVYADGQMFELSPIYHVAAD	---	---	IFLKAYGSAKRVNLEK	316			
Db	883	DKFAEMGVT	---	---	DFEMAPQY	---	---	VSSDTGGSFLDSVQNGYAFTRDYDLGI	---
Qy	317	EFQSQYQTVENMIMALISLDPYNTPMFGDSWITDKNFMAQFASWARVPFPAQAIKY	376						
Db	927	SKENKY	---	---	GTADDLVKAI	---	---	KALHSKGIKWAD	---
Qy	377	---	---	---	FATDGK	---	---	QGAEN	---
Db	979	GTPVAGSQIKNTLYVVDGKSGKQQAQYGGAFLEELQAKYFELFARKQIS	---	---	---	---	---	---	---
Qy	405	GWKNATVMVLKASPP	---	---	GEFHAQPDNGTGFELF	---	---	IKGRNFTF	---
Db	1031	GVPMDPVKIKQWSAKYFNGTNIILGRGAGYVLKQDATNTVFSLVSDNTFLPKSLVNPHG	1090						
Qy	444	---	---	---	DAGVFVY	---	---	SGDEA	---
Db	1091	TSSSVTGLVDGKGIVYTSYGNAKNAFISLGNWY	---	---	---	---	---	YFDNNGYVMTGAQS	---
Qy	485	---	---	---	RQNKWETGNLNDLVITNPSPENLDHQRSVLFINKKYFLVIDRAI	529			
Db	1142	INGANYFLSNGIQLRNAIYDNGNK	---	---	YGN	---	---	DGR	---
Qy	530	GEATGNLGVHWOLKEDSNPVPDKTRVYTYTRDGNMIMTQSLNADRTSLNEEGRKVSIV	589						
Db	1185	---	---	---	FGQWRVYFQNGIMAVGTRVHGAVQYFDAS	---	---	GFQAKGQFITTAGDKLRY	---
Qy	590	YNKELKRPVFFPKPKKNAGTQNFVSIIVPYDQKAPEISIRENKG	---	---	DFEKGLNL	645			
Db	1233	---	---	---	FDSDSNQISNRFV	---	---	RNSKGWFLFDHNGVAVTG	---
Qy	646	TLTINGKQQLVLVP	659						
Db	1266	TVTFNG	---	---	QRLYFKP	1278			

```

RESULT 12
US-09-815-242-11522
; Sequence 11522, Application US/09815242
; Patent No. US202020661569A1
; GENERAL INFORMATION:
; APPLICANT: Hasselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of

```

```

; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11522
; LENGTH: 1167
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-815-242-11522

Query Match          3.0%; Score 105.5; DB 9; Length 1167;
Best Local Similarity 19.1%; Pred. No. 1.9;
Matches 150; Conservative 107; Mismatches 298; Indels 231; Gaps 35;

```

Qy	3	TKFKRIIVFAVIALSSGNIL-----AOSSITRKDF-----DHNLEYSGLEKV	47
Db	45	TFWSQRAISFCSTQTSKNIKDSFETIMCEYDEELKXKSKFNLIKISIDHVDGTMNCKERL	104
Qy	48	NKVAAGNYDDAAKALLAYREKSKAREPD-----FSNAEKPADIQPIDKVTREMA	100
Db	105	DKLENLNTFFPNICKVLSNARCLSEGVDVPALDSVIFFDGSRAMVDIIQAVGRVMRKAKN	164
Qy	101	KALVHQFPQHKGYFYDGKDINQMWPVKONEVRWQLHRVK-----WQOMALVYHA	153
Db	165	K-----KXGYIILPIA-----LRESEIKNLDEAVKNTNFQNIWKVLKALRSHD	207
Qy	154	TG--DEKYAREWVVOYDWARKNPLGLSDNDKDFWRRELVSDRVQSILPPTFSLFVNSP-	210
Db	208	SSLVDEAIFKEKIKIFGSDDASN-----DDEELQKQKTEGSS-----NDPK	250
Qy	211	-----AFTPAFLMEFLNSYHQ-----QADYLSTHYAEOGN--HRLFEAORNLEAGVSF	256
Db	251	EAOKTLEDAILLQDLANAVNVMTKLGDRNYWENFAKKTGNIARTLNERLKELFG--KN	308
Qy	257	PEFKDSPRWQTGISVLNTEIKQVYADGMQFELSPIVHVAAIDIFLKAYGSAKRVNLEK	316
Db	309	PEIFDNFLTSLRG--NIHQSIKEBEALDMI-----ISHIITTKPIFDAIFGD-----NKN	356
Qy	317	EFPOQSVQTVENMIMALISLSLDYNTPMFGDSWITDKNPRMAQFASMARVPPANQAKY	376
Db	357	PI-----AKALDKMWIKLSDJGLEGETKDL-----KNL-----YESVKT	390
Qy	377	FATDGKGKAPNFLSKALSNAGF--YTFRSGWDK-----NATVMWLKASPP	420
Db	391	EAARAKSOKSQOELIKNLNTFFKEAFRKQSEKLGIVYTPTEVVDFFILRATNGILKKHEN	450
Qy	421	GEIHAQ-----PDNGT-----FELFIKGRNFTPD-----AGVYVYSGDBEALMKLRNW	462
Db	451	TDFNQDSQITFDPFMGTGSTFIARLLSKENDFISDEALKERFKQGLFAFD-----IVLLSY	505
Qy	463	YRQTRIHSHTLDNQNNVITKARONKWETGN-----LDVLITYNPS--YENLDHQ	511
Db	506	Y-----IALIN-----ITQAAQRDSSLKFNKNIALTDSLVDLEKNDKGIVPGFFIL	553
Qy	512	RSVLFINKKYFLVIDR-----AIGEATGNI.GVHHQMKEDSNPNVFDKTKNRVYTTYTRDGN	566
Db	554	FEDLKENKEIKTTWEEKONIRVIIENPPYSSGAKSENNDNONLHPKLEKRYVEYV--GKN	611

Qy	377	FATDGGCKAPNELSKALSNAGF--YTRFRSGWDK-:::-----NATVMVLKASPP	420
Db	391	EABARAKSQSOEELKMLNYTFKFAFRKQSEKLGIVYPIEVVDFFILRATNGIILKHFHN	450
Qy	421	GEFHAQ-----PDNGT-----FELFIKGRNFTPD-----AGVFVYSGDEAMIKLRNW	462
Db	451	TDENDQSITTFDPMGTGSGFIARLLSKENDFISDEALKERFQKGLFAPD-----IVLLSY	505
Qy	463	YRTRIHSITLTLDNQNNVITKARQNKWETGNN-----LDVLTITNPS--YPNLDHQ	511
Db	506	Y-----IALIN-----ITQAAQNRRDSSLKNFKNIALTDSLVDLEEKNDKGVIPGFVYL	553
Qy	512	RSVLFINKKYFVLIDR-----ATGEATGNI.GVHHQLKEDSNPFDKTKNRVYTYTYRDGNN	566
Db	554	FEDLKENKEIKTTWEKONIRVIIGNPPYSSGAKSENDDNNONLSHPKLEKRYVEYV--GKN	611

Qy 567 LMIQSLNADR-----TSLNEEBGKVSYYVYNKEL--KRPAFVF----- 601
Db 612 STAQNKNSRDTLHISIRMSADLKDGLGVGVVNGSFDISKSDGPRKCAVQDFSHLYA 671
Qy 602 -----EKPKK-----NAGTQNFYSIVVPYDQKAPETISIRENKGNDEPKGLN 644
Db 672 LNLRCNARTSGEERKKGDDGFDGSRVTVAIIFFVVDKADPNITFIYVEVEDYLKREAK 731
Qy 645 LTLTIN 650
Db 732 LNLLAN 737

RESULT 13
US-09-842-484A-2
; Sequence 2, Application US/09842484A
; Publication No. US20030104601A1
; GENERAL INFORMATION:
; APPLICANT: DEANGELIS, PAUL L.
; TITLE OF INVENTION: CHONDROITIN SYNTHASE GENE AND METHODS OF MAKING AND USING SAME
; FILE REFERENCE: 4605.003
; CURRENT APPLICATION NUMBER: US/09/842.484A
; CURRENT FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: 60/199,538
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 965
; TYPE: PRT
; ORGANISM: Pasteurella multocida
US-09-842-484A-2

Query Match 3.0%; Score 105; DB 11; Length 965;
Best Local Similarity 17.7%; Pred. No. 1.6;
Matches 131; Conservative 115; Mismatches 238; Indels 258; Gaps 38;

Qy 13 AVIALSSGNILAQSSSITRKDF-----DHIN-----LEYSGLEKVNKAVAAG 54
Db 95 SITGKKSENAEIRKVELVPKDFPKDLVLAPLDHVNDFTWYKNRKKSLGIGKVPVKNIGLS 154
Qy 55 -----NY-----DAAKALLAYYREKSKAREPDSNAEK 83
Db 155 IIIPTFNRSRLDITLACLNVQNTNYPFVVVADGSKENLLTIVQK-----YEQ 204
Qy 84 PADIRQPIDKVTREMAKALVHQPQPHKGYG-----FDYKGDINQWMP- 128
Db 205 KLDIK-----YVRQKDYGYQLCAVRNLGLRTAKYDFVSIILCDMAPO 246
Qy 129 -----VKDNEV-----RWQLHRVKKWQAMALVYHATGDEKYAREWYQYSDW 170
Db 247 QLWVHSYLTLELNDIVLIGPRKYVDTHNITAQFL-----NDPYLIESLPETA-- 296
Qy 171 ARKNPLGLSQDNKFWVR--PLEVSDRVQSLPPTFSLFVNSPAPTPAFLEFLNS---YH 225
Db 297 TNNPSTISKGNISLDRLEHFKTDNLRLCDSPPRYFV---AGNVAFSKEWLNVKGVWFD 353
Qy 226 QOADYLSTHYAEOGNHRLFEAQ-----RNLFAGVSP---PEFKDSPRWQTGISVLNTEIKK 279
Db 413 KVPYIYRKLPIEDSHIRIPLVSIYPAYNCA-----NYIQ---RCVDSALNQ 458
Qy 337 SLDPYNTPMFGDSWITDKNFRMAQFASWARVFPANOAIKYFATDGKQKAPNLSKALSN 396
Db 459 TVDLEVCICNDG-STDNTLEVIN-----KLYGNPNRVRIMS-----KENGGIASASN 505
Qy 397 AGFTYFRSGW-----DKRATVMVLKASPPGEHAQPDNGTLELFTKGRNFTPDA 445
Db 506 AA-VSFAKGYIGQLDSDDDYLEPDADVELCLK-----EF--LKDKTLACVYTTNRNVNPDG 557

Qy 446 GVFV--YSGDE-----AIMKLRWYRQTRIHSIL-TLDNQNMVITKARQN 487
Db 558 SLIANGYNWPEFSREKLTATTAMIAHFRMTFIRAWHLTDGFENIENIENAVDYMFLKJSEVG 617
Qy 488 KWETGNL---DVLTYNPSYPLNDHORSVLFINKKYFLVIDRAIGEATGNLGVHW--- 540
Db 618 KPHLKNKICYNRVLHGDNTSIKKLGIQK-----KNHFVVVNGSLNRQ---GINYNYND 667
Qy 541 ---QLKEDSNPVPDKTKNRVYTTYRDGNNLMIOSLNADRTSLNEEBGKVSYYVYNKELKRP 597
Db 668 KFDLDDESRYIENKT---AEYQEBMDL-----KDLK-- 697
Qy 598 AFVFERPKKNAGTQNFUSIVVP 619
Db 698 -LIQNKDAKIA-----VSIFYP 713

RESULT 14
US-10-184-485-3
; Sequence 3, Application US/10184485
; Publication No. US2003007763A1
; GENERAL INFORMATION:
; APPLICANT: THE BOARD OF REGENTS OF THE UNIVERSITY OF OKLAHOMA
; TITLE OF INVENTION: POLYMER GRAFTING BY POLYSACCHARIDE SYNTHASES
; FILE REFERENCE: 5820.551
; CURRENT APPLICATION NUMBER: US/10/184.485
; CURRENT FILING DATE: 2002-06-26
; PRIOR APPLICATION NUMBER: US/09/437,277
; PRIOR FILING DATE: 1999-11-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: WordPerfect 8.0 (saved in ASCII format)
; SEQ ID NO 3
; LENGTH: 965
; TYPE: PRT
; ORGANISM: Pasteurella multocida
US-10-184-485-3

Query Match 3.0%; Score 105; DB 15; Length 965;
Best Local Similarity 17.7%; Pred. No. 1.6;
Matches 131; Conservative 115; Mismatches 238; Indels 258; Gaps 38;

Qy 13 AVIALSSGNILAQSSSITRKDF-----DHIN-----LEYSGLEKVNKAVAAG 54
Db 95 SITGKKSENAEIRKVELVPKDFPKDLVLAPLDHVNDFTWYKNRKKSLGIGKVPVKNIGLS 154
Qy 55 -----NY-----DAAKALLAYYREKSKAREPDSNAEK 83
Db 155 IIIPTFNRSRLDITLACLNVQNTNYPFVVVADGSKENLLTIVQK-----YEQ 204
Qy 84 PADIRQPIDKVTREMAKALVHQPQPHKGYG-----FDYKGDINQWMP- 128
Db 205 KLDIK-----YVRQKDYGYQLCAVRNLGLRTAKYDFVSIILCDMAPO 246
Qy 129 -----VKDNEV-----RWQLHRVKKWQAMALVYHATGDEKYAREWYQYSDW 170
Db 247 QLWVHSYLTLELNDIVLIGPRKYVDTHNITAQFL-----NDPYLIESLPETA-- 296
Qy 171 ARKNPLGLSQDNKFWVR--PLEVSDRVQSLPPTFSLFVNSPAPTPAFLEFLNS---YH 225
Db 297 TNNPSTISKGNISLDRLEHFKTDNLRLCDSPPRYFV---AGNVAFSKEWLNVKGVWFD 353
Qy 226 QOADYLSTHYAEOGNHRLFEAQ-----RNLFAGVSP---PEFKDSPRWQTGISVLNTEIKK 279
Db 354 EEFNHWGGEDEVFG-YRLFAGKCFRVIDGGMALHQEPGKENETEBEAGKSITLKIYKE 412
Qy 280 QV---YADGMQFELSPIYHVAADIFLKAYGSAKRVNLEKFFPQSYVQTVENMTMALISI 336
Db 413 KVPYIYRKLPIEDSHIRIPLVSIYPAYNCA-----NYIQ---RCVDSALNQ 458
Qy 337 SLDPYNTPMFGDSWITDKNFRMAQFASWARVFPANOAIKYFATDGKQKAPNLSKALSN 396
Db 459 TVDLEVCICNDG-STDNTLEVIN-----KLYGNPNRVRIMS-----KENGGIASASN 505

This Page Blank (uspto)